ELSEVIER

Mathematical Biosciences 181 (2003) 17–54

Mathematical Biosciences

www.elsevier.com/locate/mbs

Modeling the impact of random screening and contact tracing in reducing the spread of HIV ☆

James M. Hyman [a], Jia Li [b,*], E. Ann Stanley [a]

[a] *Theoretical Division, MS-B284, Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

[b] *Department of Mathematical Sciences, University of Alabama in Huntsville, Madison Hall, Room 210, Huntsville, AL 35899, USA*

Received 16 November 1999; received in revised form 4 June 2002; accepted 27 June 2002

Abstract

Mathematical models can help predict the effectiveness of control measures on the spread of HIV and other sexually transmitted diseases (STDs) by reducing the uncertainty in assessing the impact of intervention strategies such as random screening and contact tracing. Even though contact tracing is one of the most effective methods used for controlling treatable STDs, it is still a controversial strategy for controlling HIV because of cost and confidentiality issues. To help estimate the effectiveness of these control measures, we formulate two models with random screening and contact tracing based on the differential infectivity (DI) model and the staged-progression (SP) model. We derive formulas for the reproductive numbers and the endemic equilibria and compare the impact that random screening and contact tracing have in slowing the epidemic in the two models. In the DI model the infected population is divided into groups according to their infectiousness, and HIV is largely spread by a small, highly infectious, group of superspreaders. In this model contact tracing is an effective approach to identifying the superspreaders and has a large effect in slowing the epidemic. In the SP model every infected individual goes through a series of infection stages and the virus is primarily spread by individuals in an initial highly infectious stage or in the late stages of the disease. In this model random screening is more effective than for the DI model, and contact tracing is less effective. Thus the effectiveness of the intervention strategy strongly depends on the underlying etiology of the disease transmission.

☆ This research was supported by the Department of Energy under contracts W-7405-ENG-36 and the ASCR Applied Mathematical Sciences Program KC-07-01-01.

* Corresponding author. Tel.: +1-256 890 6470; fax: +1-256 895 6173.

E-mail address: lij@email.uah.edu (J. Li).

0025-5564/03/$ - see front matter
PII: S0025-5564(02)00128-1

Keywords: AIDS; Mathematical modeling; Epidemic modeling; Screening; Contact tracing; Reproductive number; Endemic equilibrium; Sensitivity

1. Introduction

Mathematical models based on the underlying transmission mechanisms of the disease can help the medical/scientific community understand and anticipate the spread of an epidemic and evaluate the potential effectiveness of different approaches for bringing an epidemic under control. Models can be used to improve our understanding of the essential relationships between the social and biological mechanisms that influence the spread of a disease. The relative influence of various factors on the spread of the epidemic, as well as the sensitivity to parameter variation, can be ascertained. Because the transmission dynamics form a complex non-linear dynamical system, the behavior of the epidemic is a highly non-linear function of the parameter values and levels of intervention strategies. This at times may even lead to changes in infection spread that are counter to both intuition and simple extrapolated predictions. We can use the knowledge gained from studying models to help set priorities in research, saving time, resources, and lives.

Screening is one of the most common strategies used to control the spread of HIV infection. State health services provide anonymous or confidential screening to individuals who come in on a voluntary basis, perhaps because they think they may have been exposed to HIV, or they are part of a high risk group. Pregnant women are often screened for HIV infection. Infected individuals are also identified when they donate blood, draw blood as part of a physical exam, or are tested for HIV for other reasons. Models can be used to study the impact of such screening programs. They can also be applied to study more costly contact tracing programs.

Contact tracing, also known as 'partner notification by provider referral', is one of the most effective strategies for controlling treatable sexually transmitted diseases (STDs) such as syphilis and gonorrhea. These programs ask infected individuals to identify other people whom they may have infected or been infected by. Trained personnel then attempt to contact the named partners, inform them that they had an infected partner, educate them, and provide them with opportunities to be tested for the infection. If they are infected, they can begin treatment and stop unknowingly spreading infection.

Although contact tracing has been used for years as an effective method for controlling curable STDs, it remains controversial and hotly-debated as a strategy for controlling HIV. The advantages of identifying partners of those infected with HIV are not as clear as they are with easily treated infections. However, the gravity of HIV infection and the magnitude of the epidemic make it imperative that we understand the relative effectiveness of all possible control approaches.

Confidentiality issues, the cost of the program, and the likelihood that fewer people will come in for testing are important considerations when deciding whether or not to implement contact-tracing. Some specialists in the field argue that the potential for putting people at serious risk of ostracization and even physical harm from others are not worth the potential gain. People are less likely to voluntarily be tested when they are asked, or even required by law, to name their sexual partners. This is of particular concern when there is a possibility of domestic violence [1,26,31]. Until recently, very little could be done for HIV-infected people, and thus informing them of their infection was like handing them a death sentence. Many health service workers were reluctant to do this.

Other specialists in the field have argued that contact tracing is more effective than screening programs, which often attract mostly the 'worried well' who are not at high risk [5,18]. It is also argued that the rights of those who have been exposed to know about their exposure, and the need to stop the chain of infection, should supersede the rights of the infected to privacy [28]. The fact that many studies have found that contact tracing is an effective strategy for finding and counselling infected people [14,21,29,31] lends force to their argument. Another argument in favor of contact tracing is that it can "delineate the risk networks hosting transmission and provide empiric estimates for mathematical model parameters" [23].

With today's new treatments for HIV infection, some of the earlier arguments against contact tracing have been partly eliminated. Although concerns still remain about decreased participation in testing, and domestic violence, there are more and more reasons to identify infected people as early in the course of infection as possible, to allow them to be promptly treated and to reduce the chance that they will unknowingly transmit the disease.

While it seems likely that contact tracing could be as effective in controlling the spread of HIV as it has been for other STDs, there are few analytical studies to estimate what fraction of the population should be screened, what fraction of their partners should be contacted in order for the program to have a significant effect on the spread of the epidemic, or how much the behavior of this tested population needs to change. Scientists are beginning to develop models to study these questions. Kretzchmar et al. [13] used simulations of the spread of gonorrhea and chlamydia to study random screening and contact tracing, finding that, for their model, treatment of even a small fraction of the partners of those with symptoms could completely halt the epidemic, whereas screening of even large fractions of the population had little effect. Their model neglected 'snowballing', the situation where not only the partners of the originally screened infecteds, but also the partners of those partners, and so on, are traced, until no more infected individuals are found. They also neglected the situation where a past partner of an infected individual was infected by someone else either before or after their partnership. Müller et al. [20] incorporated snowballing and infection of partners by others, and analytically studied contact tracing and screening in a stochastic model of a simple SIRS (susceptible-infected-removed-susceptible) epidemic in a population of fixed size. They derived formulas for the reproductive number under different assumptions for the stochastic model, and created a deterministic model with the same reproductive number.

Here we use a different methodology to develop two models for HIV spread which include contact tracing and random screening in populations with variable sizes. We develop the models directly as differential equations, using approximations to estimate terms in our equations, rather than attempting to derive them from a stochastic or simulation model. Differential equations allow us to quickly obtain insights into the dynamics of the two models. As in [13], we neglect snowballing, but, unlike [13], we do account for the possibility that partners of infecteds were infected by someone other than the index case.

These models are extensions of the two models developed in detail in [8,9]. We have chosen them specifically to address questions about whether or not contact tracing can be effective, given that viral loads vary so much between individuals and within individuals over the course of their infection. The differential infectivity (DI) model divides the infected population into groups according to their infectiousness, and accounts for differences in rates of developing AIDS. In contrast with the DI model, we also studied a simple version of a staged-progression (SP) model,

in which every infected individual goes through the same series of stages. The parameters we use for the SP model give a short, early, highly infectious, stage equivalent to the acute phase of infection; a middle period of low infectiousness; and a late chronic stage with higher infectiousness. Thus the DI model captures individual differences and the SP model captures differences in time within the same individual.

In [8,9] we simulated the transient dynamics and studied the sensitivity of both models using parameters derived from the literature. We also developed a robust method for initializing multigroup epidemic models. For the SP model, these studies provided further insight into the observations in [11,12] that, when partner acquisition rates are high, the bulk of the infections early in the epidemic are caused by those in the acute infectious stage. For the DI model, we showed that a small number of individuals who are highly infectious during the chronic stage have a disproportionate impact on the epidemic, even though they have a short life expectancy. Both models were found to be very sensitive to the probability of transmission per contact and the sexually active removal rate.

In this paper we first review the mathematical formulation of the original DI and SP models, and then add terms to account for random screening and contact tracing. In developing these new terms, we carefully justify our assumptions. We find reproductive numbers for both models, and show that they have a unique endemic equilibrium which exists if and only if the epidemics are above threshold. Then we analyze the models to assess the impact of intervention strategies. We use numerical simulations to compare the impact of the strategies on the epidemic, and use our analytical formulas for the reproductive numbers and the endemic equilibria to examine in more detail the sensitivity of both models to the level of intervention strategy. Screening and changing the behavior of 5% of the high-risk population every year significantly slows the epidemic for both models, reducing the number of infections by more than a third. For the DI model, adding contact tracing to the screening is an effective approach to identifying the superspreaders and further slows the disease spread by a significant amount: when half of all partners can be found, it drops the number of people infected well below half the number who would get infected with no controls. For the SP model, contact tracing also drops the number of infections significantly, but not as much as for the DI model. If the SP model holds, then it appears that contact tracing might primarily identify individuals after they are past the most infectious stage, and it is possible that public health might not be served by an expensive contact tracing program. However, if the DI model is closer to the underlying disease etiology, then the epidemic can be significantly slowed if the superspreader group can be identified and removed from the transmission network.

In deriving our models, we find that two of the factors we neglect are difficult to justify. We finish this paper by estimating the size of one of these terms, and showing that it is small compared to the terms we accounted for in the model. Then we give a formula for the other factor, and arguments as to why it is reasonable to neglect it as well.

2. The differential infectivity and staged-progression models

Here we briefly describe the DI and SP models without random screening or contact tracing and review the analysis for R_0 and the endemic equilibrium [8,9]. The intervention strategies will be added to these basic models in the next section.

2.1. The differential infectivity model

During the chronic stage of infection, viral levels differ by orders of magnitude between individuals. Those with high viral loads in the chronic phase tend to progress rapidly to AIDS, while those with low loads tend to progress slowly to AIDS [3,4,22,30]. The DI model accounts for the distribution of times from infection to AIDS by assuming variations between individuals in their duration of infection, dividing the infected population into n groups.

The equations for the DI model illustrated in Fig. 1 are

$$\begin{aligned}
\frac{dS}{dt} &= \mu(S^0 - S) - \lambda S,\\
\frac{dI_i}{dt} &= p_i\lambda S - (\mu + \nu_i)I_i, \quad i = 1,\ldots,n,\\
\frac{dA}{dt} &= \sum_{j=1}^{n} \nu_j I_j - \delta A,\\
\lambda(t) &= \sum_{i=1}^{n} \lambda_i(t), \quad \lambda_i(t) = r\beta_i \frac{I_i(t)}{N(t)},
\end{aligned} \tag{2.1}$$

where $N(t) = S(t) + \sum_{j=1}^{n} I_j(t)$. Here S denotes the susceptibles, I_i denotes the number of infected individuals in group i, and A denotes the number of infected individuals no longer transmitting the disease. S^0 is the constant steady state population maintained by the inflow and outflow when no virus is present in the population. The total removal rate μ accounts for both natural death in the absence of HIV infection and people moving in and out of the sexually active susceptible population due to behavior changes or physical migration. $\lambda(t)$ is the rate of infection per

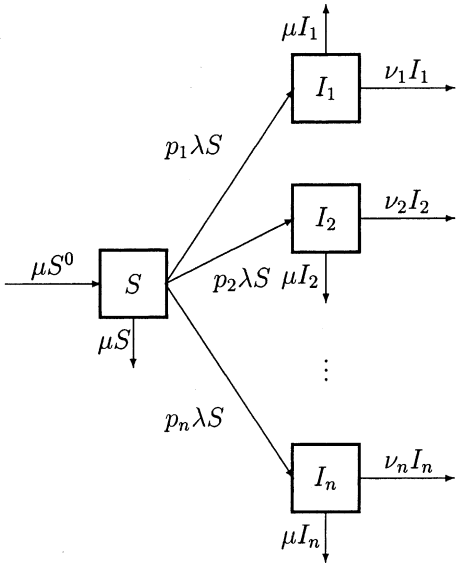

Fig. 1. The DI model divides the infected population into groups according to their infectiousness or differences in rates of developing AIDS. In this model HIV is primarily spread by a small, highly infectious, group of superspreaders.

susceptible, r is the partner acquisition rate, and β_i is the probability of transmission per partner from infected individuals in group i. Upon infection, an individual enters subgroup i with probability p_i, where $\sum_{i=1}^{n} p_i = 1$, and stays in this group until becoming inactive in transmission. Finally, v_i is the rate at which infected individuals in group i enter group A, and δ is the death rate of people in group A. All infected individuals are assumed to eventually enter group A prior to death due to their infection.

2.2. The staged-progression model

The viral burden during HIV infection varies as a function of time within an individual. Initially, the HIV-1 RNA levels in plasma and serum can become extremely high during the first weeks of acute primary infection, even before there is a detectable immune response [24,25]. These levels are higher than at any other time during infection. Acute primary infection is followed by a chronic phase during which the HIV RNA levels drop several orders of magnitude and remain at a nearly constant level for years [7,22,30]. In the late chronic stages of an infection, the HIV-1 RNA levels may increase as much as ten-fold [7] over what they have been during the rest of the chronic stage. The SP model accounts for the temporal changes in the infectiousness of an individual by a staged Markov process of n infected stages, progressing from the initial infection to AIDS.

The equations for the SP model illustrated in Fig. 2 are

$$\begin{aligned}
\frac{dS}{dt} &= \mu(S^0 - S) - \lambda S,\\
\frac{dI_1}{dt} &= \lambda S - (\gamma_1 + \mu)I_1,\\
\frac{dI_i}{dt} &= \gamma_{i-1}I_{i-1} - (\gamma_i + \mu)I_i, \quad 2 \leqslant i \leqslant n,\\
\frac{dA}{dt} &= \gamma_n I_n - \delta A,\\
\lambda(t) &= \sum_{i=1}^{n} \lambda_i(t), \quad \lambda_i(t) = r\beta_i \frac{I_i(t)}{N(t)},
\end{aligned} \tag{2.2}$$

where now I_i is the number of infected individuals in each infected stage. Note that all individuals go into group 1 upon infection. γ_i is the rate at which individuals move from stage i of infection to stage $i+1$. The meanings of S^0, μ, r, and δ are the same as in the DI model, and β_i is the probability of transmission per partner from infected individuals in stage i. Previous studies of SP models can be found in [2,10–12,15–17].

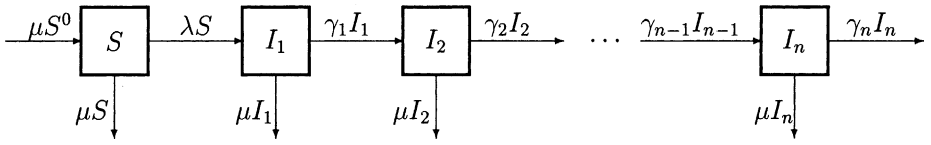

Fig. 2. In the SP model every infected individual goes through the same series of stages. This model can account for a short early highly infectious stage equivalent to the acute phase of infection, a middle period of low infectiousness, and a late chronic stage with higher infectiousness.

2.3. Transmission probability

The parameter r enters the model both as a multiplicative factor and through the dependence of the transmission probabilities per partner, β_i, on the average number of contacts per partner, c, which in turn depends on the number of contacts per partner ($c = c(r)$).

If ζ_i is the transmission probability per contact in group i, the probability that a susceptible individual will not be infected by a single contact with an infected individual is $1 - \zeta_i$. Hence the probability that a susceptible individual will avoid infection when they have $c(r)$ contacts with an infected partner is $(1 - \zeta_i)^{c(r)}$, and the probability of transmission per partner from an infected person in group i is

$$\beta_i = 1 - (1 - \zeta_i)^{c(r)}. \tag{2.3}$$

Our choice for $c(r) = 104r^{-\eta} + 1$ in Section 5 gives approximately two contacts per week for people with one partner per year, and decreases to about one contact per partner as r gets large [9]. The parameter η controls how fast this function decreases. In the simulations presented in Section 5, we set $\eta = 1$.

Let $\bar{\tau}_i$ be the mean duration of infection in group i. Then, for the DI model, $\bar{\tau}_i = 1/(\mu + \nu_i)$, and for the SP model, $\bar{\tau}_i = 1/(\mu + \gamma_i)$. The mean duration of infection for the whole population for the DI model and SP model are given by $\bar{\tau} = \sum_{i=1}^{n} p_i \bar{\tau}_i$ and $\sum_{i=1}^{n} q_i \bar{\tau}_i$, respectively, where q_i is defined as in Table 1. Based on these notations, the mean transmission probability per contact $\bar{\zeta}$ for the DI and SP models are

$$\bar{\zeta}^{\mathrm{D}} = \sum_{i=1}^{n} p_i \frac{\bar{\tau}_i}{\bar{\tau}} \zeta_i, \quad \bar{\zeta}^{\mathrm{S}} = \sum_{i=1}^{n} q_i \frac{\bar{\tau}_i}{\bar{\tau}} \zeta_i. \tag{2.4}$$

Table 1
Reproductive number R_0, mean duration of infection in group i, $\bar{\tau}_i$, mean duration of infection for the whole population $\bar{\tau}$, mean transmission probability $\bar{\beta}$, equilibrium infection rate λ^*, susceptible population S^*, equilibrium infected group population I_i^*, equilibrium total infected population I_T^*, and equilibrium relative impact ρ_i^* for both models

Name	DI Model	SP Model
R_0	$r\bar{\tau}\bar{\beta}$	$r\bar{\tau}\bar{\beta}$
$\bar{\tau}_i$	$\frac{1}{\mu+\nu_i}$	$\dfrac{1}{\mu+\gamma_i}$
$\bar{\tau}$	$\sum_{i=1}^{n} p_i\bar{\tau}_i$	$\sum_{i=1}^{n} q_i\bar{\tau}_i$
$\bar{\beta}$	$\sum_{i=1}^{n} p_i\beta_i\bar{\tau}_i/\bar{\tau}$	$\sum_{i=1}^{n} q_i\beta_i\bar{\tau}_i/\bar{\tau}$
q_i	Undefined	$\prod_{j=1}^{i-1}\gamma_j\bar{\tau}_j$
S^*	$\frac{\mu S^0}{\mu+\lambda^*}$	$\frac{\mu S^0}{\mu+\lambda^*}$
I_i^*	$p_i\bar{\tau}_i S^*\lambda^*$	$q_i\bar{\tau}_i S^*\lambda^*$
I_T^*	$S^*(R_0-1)$	$S^*(R_0-1)$
λ^*	$\dfrac{R_0-1}{\bar{\tau}}$	$\dfrac{R_0-1}{\bar{\tau}}$
ρ_i^*	$\dfrac{p_i\beta_i\bar{\tau}_i}{\bar{\beta}\bar{\tau}}$	$\dfrac{q_i\beta_i\bar{\tau}_i}{\bar{\beta}\bar{\tau}}$

2.4. The reproductive number and endemic equilibrium

We proved in [8] that both of these models have two equilibria: the infection-free equilibrium (given by $S = S^0, I_i = 0$), and the endemic equilibrium (given by $S = S^* > 0, I_i = I_i^* > 0$). The endemic equilibrium is the asymptotic distribution of the infection in the population once the initial transients have settled down. Analyzing the stability of the infection-free equilibrium gives the reproductive number, which specifies the conditions under which the number of HIV infected individuals will initially increase or decrease when there are a small number of them at the start. The reproductive number, R_0 is defined such that if $R_0 < 1$ the modeled epidemic dies out and if $R_0 > 1$ the epidemic spreads [6]. The reproductive number is obtained by investigating the stability of the infection-free equilibrium at which the components of infected groups are zero. If $R_0 < 1$, this infection-free equilibrium is the unique equilibrium. If $R_0 > 1$, the infection-free equilibrium becomes unstable and there appears, for both models, a unique endemic equilibrium at which the components of infected groups are positive.

The reproductive number can be written

$$R_0 = r\bar{\tau}\bar{\beta} \tag{2.5}$$

for both models. Here $\bar{\tau}$ is the mean duration of infection, and $\bar{\beta}$ is the mean probability of transmission per partner. We also found formulas for the endemic equilibrium, and proved that there exists a non-trivial equilibrium if and only if the reproductive number R_0 is greater than 1. If the endemic equilibrium exists, it is always locally asymptotically stable. The formulas for all of these quantities are given in Table 1.

The relative importance of each infection group in maintaining the chain of transmission is measured by the relative fraction of individuals being infected by each group. The *relative impact* of I_i on the rate of infection is

$$\rho_i(t) = \frac{\lambda_i(t)}{\lambda(t)} = \frac{\beta_i I_i(t)}{\sum_{j=1}^{n} \beta_j I_j(t)}. \tag{2.6}$$

Note that the formulas for the DI and SP models in Table 1 have the same form, with p_i and v_i from the DI model being replaced by q_i and γ_i for the SP model formulas. However, while it could be argued that v_i and γ_i are both progression rates and thus play similar roles in both models, q_i is quite different from p_i. Not only is q_i a derivative quantity, but also $q_1 = 1$ so that the sum of the q_i is larger than one, while the p_i sum to one. The similarity of formulas can be deceptive in making the models appear more similar than they are.

3. Random screening and contact tracing models

In this section we modify the DI and SP models to account for two types of control programs. The first type, random screening, tests broad sectors of the population for HIV infection. Random screening programs include the screening and notification of blood donors and pregnant women, and anonymous or confidential testing sites. People come to these sites somewhat at random, either to donate blood or because they believe they may be at risk for HIV infection. In all of these cases, when people are identified as infected, they are counselled about risk behaviors. We assume that these programs test and counsel the population at a rate ε.

Once infected people who know of their infection status have gone through a counselling program, they have a wide variety of reactions. Ideally, all of them would either abstain from sex, or use condoms with all partners. However, unfortunately, this has not been found to be the case [19]. Some people change behaviors dramatically and some do not change much at all. Accounting for the many nuances of behavior change, such as a decrease in the number of partners versus a shift to condom use, is beyond the scope of this model. We assume that a fraction, κ, of the counselled population leaves the high risk population, and the behavior of the remaining fraction, $1-\kappa$, remains unchanged. Because $\sigma := \kappa\varepsilon$ is small, we neglect the fact that those who have already been tested by random screening are unlikely to be tested again, and lump these people back in with the general infected population.

We assume that the rate, ε, that someone is identified as infected by random sampling, and the fraction, κ, of these people who change their behavior are homogeneous in the population and remain constant over time. Thus we subtract a term $\kappa\varepsilon I_i = \sigma I_i$ from the equation for the infected group, I_i, and add it into the equation for a new group, I_{C_i}, the tested and counselled infected people who have changed behavior. Because some of the partners of the infected and screened people will also become part of I_{C_i}, for clarity in what follows we refer to the infected people found via screening as the *screened infecteds*.

The second type of program we model is active contact tracing. These HIV-control programs operate on top of screening programs. When infected people have been identified by a screening program, they are asked to identify their partners for the past T_A years, where, in most programs described in the literature, T_A is between six months and a year. A fraction f of those past partners are named, located, tested for HIV infection, and counselled.

In this initial model we neglect 'snowballing'. If we call people who are named by screened infecteds, tested and found to also be infected, *level two traced infecteds*, then snowballing occurs when level two traced infecteds are asked to name partners, and those partners are traced and tested. We can call the people who have been found to be infected because they were contacts of level two traced infecteds *level three traced infecteds*. With snowballing, partners of level three traced infecteds are also traced, possibly yielding some level four traced infecteds, and the chain is followed until no more infected people are found. Thus, by neglecting snowballing, we do not account for traced infecteds at level three and beyond.

We justify this because the data reported in the literature seems to indicate that the number of infected people found through snowballing in the typical contact-tracing program is small compared to the number of infected people found who are direct partners of screened infecteds. For example, in [14], only 46% of those eligible to participate in the study agreed to do so, and named partners. Only half of their named partners were located, implying that at most 23% of eligible people's partners were contacted. If this contacted group were similar to those participating in the study, about 46% of them would agree to be tested, implying that about 11% of level two partners would be tested. Because infection levels are usually less than 50% of any population, most of the ones who did agree to be tested would not be found to be infected. If we have a population which averages five partners over the period in question, we are thus talking about finding at most (at 50% infection rates) around one level two traced infected for every four screened infecteds.

Snowballing occurs when the contact tracing program goes to the next level: infected people from level two name partners and some of them are also found to be infected. The impact of these terms is multiplicative and multiplying the small factors together results in an even smaller effect.

When the snowballing is expanded to level three, there is only an average of one level three traced infected for every sixteen screened infecteds. Thus, in this situation, including snowballing would not dramatically change our predictions.

This would not be the case if we wished to model aggressive contact tracing programs, where most of the people traced through networks are located, such as in the program described in [31]. In such a situation, the multiplicative factors become larger and snowballing can be an important factor. The models we consider could be easily modified to account for a small snowballing effect by assuming that the same fraction of partners of the identified infected partner are infected as in the original index case. We have not included this factor in the current model, and therefore our estimates on the impact of contact tracing slightly underestimate the full impact of a comprehensive program.

The rate that the active contact tracing program identifies infected people who were sexual contacts of screened infecteds from group i is the rate that infecteds are screened, $\varepsilon I_i(t)$, times the fraction of their partners who can be named, located, and tested, f, times the number of partners that they had during the time period $(t - T_M, t)$ who are infected at time t, C_i. Here we define T_M to be the minimum of the time period T_A that they are asked about, and the time period for which they can recall information such as names and how to locate them, since this is a highly active population, where individuals may not be able to give accurate information about partners for very long periods of time.

We assume that none of these partners have already been identified as being infected or have left the population due to death or AIDS. Then, since the same fraction, κ, of these identified infecteds will change behaviors as in the screening program, we remove infecteds from the population at the increased rate $f\sigma C_i$. This assumption is valid when σ, μ, and the rates of developing

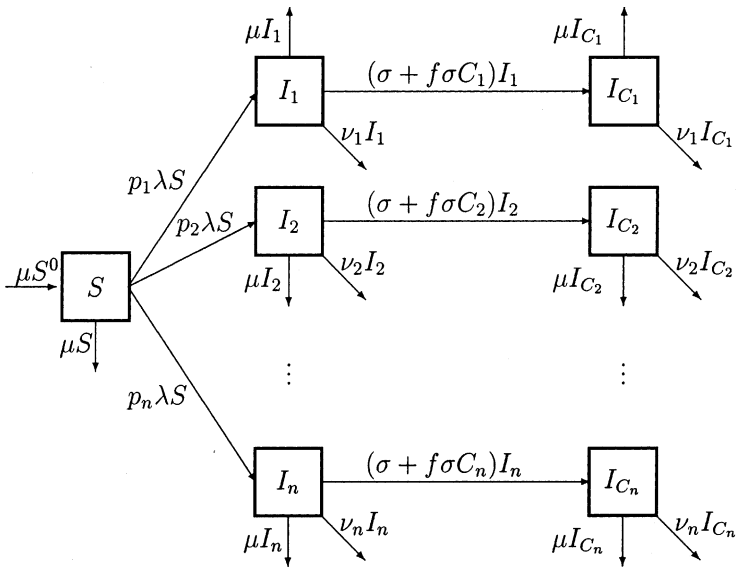

Fig. 3. The SCT-DI model with random screening and contact tracing differs from the original DI model in Fig. 1 in that it includes a new category of infected individuals, I_{C_i}, who have been identified as infected and are no longer spreading the virus.

AIDS are all small, so that the possibility the partner has left the high-risk population before being located is small.

The number of infected partners, C_i, is the sum of three terms: L_i, M_i, and O_i. L_i is the average number of people a screened infected, who was in group I_i at the time of screening, contacted in the past T_M years who were already infected before the contact. M_i is the average number of partners of this screened infected in the past T_M years who (1) were not infected at the time of their contact and who (2) were infected through this contact. O_i is the average number of partners of the screened infected in this time period who (1) were not infected at the time of the contact; (2) were not infected by the screened infected; and (3) became infected by time t.

We estimate these three terms by assuming T_M is small compared to μ^{-1}, either because activity levels are high, and therefore people can only identify their past partners and provide contact information (such as phone numbers) for a short period of time, or because they are not asked about a long time period. For example, the index cases in [14] were asked to name partners in the past year, which is short compared to the average time μ^{-1} a person stays in the high risk group.

The equations for the random screening and contact tracing DI model (SCT-DI model) illustrated in Fig. 3 are

$$\begin{aligned}
\frac{dS}{dt} &= \mu(S^0 - S) - \lambda S,\\
\frac{dI_i}{dt} &= p_i\lambda S - (\mu + v_i + \sigma + f\sigma C_i)I_i, \quad i = 1,\dots,n,\\
\frac{dI_{C_i}}{dt} &= -(\mu + v_i)I_{C_i} + (\sigma + f\sigma C_i)I_i, \quad i = 1,\dots,n,\\
\lambda(t) &= \sum_{i=1}^{n}\lambda_i(t) = \sum_{i=1}^{n} r\beta_i\frac{I_i(t)}{N(t)},
\end{aligned} \tag{3.1}$$

where $N(t) = S(t) + I(t)$, $I(t) = \sum_{i=1}^{n} I_i(t)$, does not include the identified infected people, and $C_i(t) = L_i(t) + M_i(t) + O_i(t)$. We leave out the equation for the A group because we assume that they are no longer active and hence play no role in the transmission dynamics of HIV in the model.

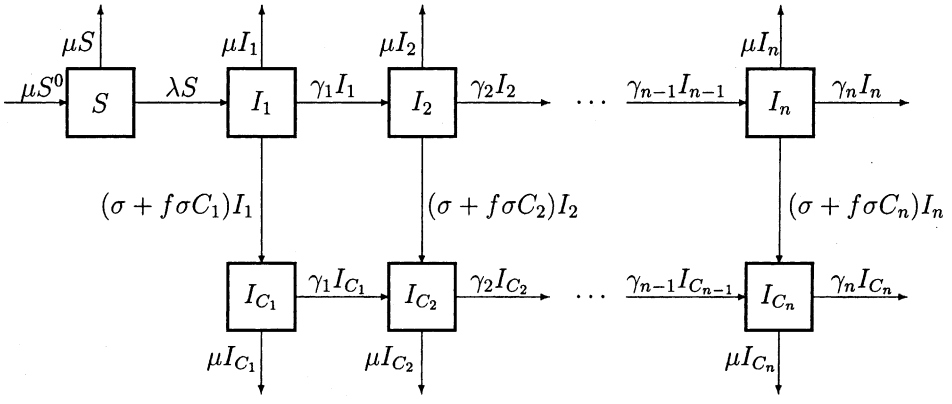

Fig. 4. The SCT-SP model with random screening and contact tracing differs from the original SP model in Fig. 2 in that it includes a new category of infected individuals who have been identified as infected and are no longer spreading the virus.

The equations for the random screening and contact tracing SP model (SCT-SP Model) illustrated in Fig. 4 are

$$\begin{aligned}
\frac{dS}{dt} &= \mu(S^0 - S) - \lambda S,\\
\frac{dI_1}{dt} &= \lambda S - (\gamma_1 + \mu + \sigma + f\sigma C_i)I_1,\\
\frac{dI_i}{dt} &= \gamma_{i-1}I_{i-1} - (\gamma_i + \mu + \sigma + f\sigma C_i)I_i, \quad 2 \leqslant i \leqslant n,\\
\frac{dI_{C_1}}{dt} &= -(\gamma_1 + \mu)I_{C_1} + (\sigma + f\sigma C_1)I_1,\\
\frac{dI_{C_i}}{dt} &= \gamma_{i-1}I_{C_{i-1}} - (\gamma_i + \mu)I_{C_i} + (\sigma + f\sigma C_i)I_i, \quad 2 \leqslant i \leqslant n,\\
\lambda(t) &= \sum_{i=1}^{n} \lambda_i(t) = \sum_{i=1}^{n} r\beta_i \frac{I_i(t)}{N(t)},
\end{aligned} \tag{3.2}$$

where $N(t) = S(t) + I(t)$, $I(t) = \sum_{i=1}^{n} I_i(t)$, does not include the identified infected people, and $C_i(t) = L_i(t) + M_i(t) + O_i(t)$. We once again leave out the equation for A.

Notice that in both models the total number of infected people become the total number of *unidentified* infected people, $I(t) = \sum_{i=1}^{n} I_i(t)$, and that the total active population now is $N(t) = S(t) + I(t)$, with I_{C_i} removed.

3.1. Estimation of $L_i(t)$

Next we estimate the average number of previously infected partners of a screened infected, L_i, for both models. Let $T_{i,\mathrm{tot}}(t)$ be the average number of years that an infected person from group i has been in the high risk population at time t. Because we assume that a person cannot have any contacts with an infected person before they enter the high risk population, a screened infected can name partners only for whichever is shorter: the time they have been in the high risk population, or T_{M}. Let $\widehat{T}_i(t) = \min\{T_{\mathrm{M}}, T_{i,\mathrm{tot}}(t)\}$. Then the average number of previously infected partners this individual has had in the past T_{M} years is

$$L_i(t) = r \int_{t-\widehat{T}_i(t)}^{t} \frac{I(s)}{N(s)}\,ds. \tag{3.3}$$

To calculate this quantity exactly requires that we can estimate $T_{i,\mathrm{tot}}(t)$, and the integral of I/N. Estimating $T_{i,\mathrm{tot}}(t)$ also requires an integral over the past. However, integrals over the past greatly complicate a differential equation model, turning it into an integro-differential equation model. To avoid this, we observe that if the spread of HIV is not extremely rapid, an infected person will have been in the high risk population for quite a while prior to infection. With this assumption, we observe that, since $T_{i,\mathrm{tot}}(t)$ is the sum of the time spent as a susceptible and the time spent as an infected, it is close in magnitude to $1/\mu$. Because we have assumed that T_{M} is small compared to $1/\mu$, then $\widehat{T}_i(t) = T_{\mathrm{M}}$ for all i.

When HIV is not spreading extremely rapidly, then for small T_{M}, the above integral for L_i is over a relatively small time interval, $(t - T_{\mathrm{M}}, t)$. It therefore is reasonable to approximate $N(s)$ and

$I(s)$ by their values at time t during this small time interval. Thus we estimate $L_i(t)$ for both models as

$$L_i(t) = \frac{rT_M I(t)}{N(t)}. \tag{3.4}$$

Note that $L_i(t)$ is independent of i. This approximation greatly simplifies both the models and the calculation of their reproductive numbers.

3.2. Estimation of $M_i(t)$

The procedure for estimating the average number of partners infected by the identified, infected individual, M_i, is different for the SCT-DI and SCT-SP models.

3.2.1. $M_i(t)$ for the SCT-DI model

Let $T_i(t)$ be the mean time that an infected person in group I_i has been in that group. We approximate this as one over the rate at which people leave group i, $\mu + \nu_i + \sigma + f\sigma C_i(t)$. This estimate will be most appropriate when the infected population is changing slowly, so that the mean time is close to the inverse rate. (Note that if this rate were a constant, this inverse would be both the mean time that people stay in group I_i as well as the mean time that people in group I_i have been in the group when the population is at equilibrium.) In order to develop a model for which we can find an equilibrium, we approximate $C_i(t)$ in our estimate of $T_i(t)$ by its value at the infection-free equilibrium, C_i^0:

$$T_i(t) \approx \bar{\tau}_i^0 = \frac{1}{\mu + \nu_i + \sigma + f\sigma C_i^0}. \tag{3.5}$$

Note that an exact expression for T_i would require adding another variable, the time since entering group i, to our model. This would lead us to a set of integro-partial differential equations. Not only would that greatly complicate the mathematics involved, it would also require the specification of a distribution of the population with duration of infection at the initial time. Because T_i early in the simulations would depend entirely upon this initial distribution, these initial conditions could introduce unintended effects. While setting T_i to its equilibrium value introduces a small error in the early stages of the epidemic, it greatly simplifies the model formulation and analysis.

Let $\widetilde{T}_i(t) = \min\{T_M, T_i(t)\}$. Then the average number of people that this infected person has infected is

$$M_i(t) = r\beta_i \int_{t-\widetilde{T}_i(t)}^{t} \frac{S(s)}{N(s)} \mathrm{d}s.$$

As above, if we make the simplifying assumption that S and N have been at their values of time t for the length of time $\widetilde{T}_i(t)$, then for the SCT-DI model,

$$M_i(t) \approx \frac{r\widetilde{T}_i\beta_i S(t)}{N(t)}. \tag{3.6}$$

Defining $\alpha_i = \beta_i \widetilde{T}_i$ allows us to split this into a part which depends upon i and is independent of time and a part which is independent of i and dependent on time:

$$M_i(t) \approx \frac{rS(t)}{N(t)} \alpha_i. \tag{3.7}$$

3.2.2. $M_i(t)$ for the SCT-SP model

For the SCT-SP model, an exact expression for M_i is given by integrating over the past as above with the SCT-DI model, except that the value for β_i now depends on the group the infected was in at the past time, s. Writing this as $\beta(s)$, we have

$$M_i(t) = r \int_{t-\widetilde{T}_i(t)}^{t} \frac{\beta(s)S(s)}{N(s)} \mathrm{d}s.$$

As before, we approximate the populations by their time t values, and the mean time a person who is in group I_i has been in that group, T_i, by

$$T_i \approx \bar{\tau}_i^0 = \frac{1}{\mu + \gamma_i + \sigma + f\sigma C_i^0}. \tag{3.8}$$

Then we can write $M_i(t)$ in the same form as for the DI model:

$$M_i(t) = \frac{rS(t)}{N(t)} \alpha_i,$$

where

$$\alpha_i = \int_{t-\widetilde{T}_i(t)}^{t} \beta(s)\, \mathrm{d}s.$$

The estimation of α_i is different for different groups. For group 1, people have the same β_i for the whole time they have been infected. Thus

$$\alpha_1 = \beta_1 \widetilde{T}_1,$$

where $\widetilde{T}_1 = \min\{T_{\mathrm{M}}, \bar{\tau}_1^0\}$.

For the remaining groups ($i > 1$), the infectivity varies over the duration of their infection. Because we assume that people can identify a fraction of their partners for the past T_{M} years, we can convert this time for people in the infected group I_i to the index $J(i)$ of the earliest infected group that an infected person in I_i was in when they may have infected another person, where $i > 1$ and $J(i) \leqslant i$. That is, a person in I_i can identify past partners while they were in groups I_j where $j \in [J(i), i]$. For example, if i is 3 and $J(3) = 2$, people in group I_3 can identify partners from the time when they were in group I_2, but they cannot identify partners from the times prior to entering group I_2.

Define $T_{k,\mathrm{inf}}$ to be the average length of time period that people in group I_k have been infected, and T_k^* to be the average length of time period that those people entering group I_{k+1} from I_k have been infected. Because these people have survived to the kth group and are still in the active population, we do not include the removal rate, $\mu + \sigma$, when estimating T_k^*. Defining

$$g_{i,j} = \sum_{l=i}^{j} \frac{1}{\gamma_l}, \quad \text{and} \quad G_{j,i} = \bar{\tau}_i^0 + g_{j,i-1},$$

we approximate

$$T_k^* = g_{1,k},$$
$$T_{k,\inf} = G_{1,k}.$$

The index $J(i)$ is determined by T_M and $T_{i,\inf}$. That is, $J(i)$ is the index of the group for which

$$T_{i,\inf} - T^*_{J(i)} < T_M \leqslant T_{i,\inf} - T^*_{J(i)-1},$$

or more specifically,

$$G_{J(i)+1,i} < T_M \leqslant G_{J(i),i}. \tag{3.9}$$

There are three possible cases.

Case 1. $J(i) = i$ and $T_M \leqslant \bar{\tau}_i^0$.

In this case, the average infected person arrived in their current infected group so long ago that they cannot identify partners they had while they were in a previous group. For this case we use the estimate

$$\alpha_i \approx T_M \beta_i. \tag{3.10}$$

Case 2. $1 \leqslant J(i) < i$.

In this case, T_M is longer than the time people have been in group I_i, but shorter than the time they have been infected. The average time they have been in group I_i is $\bar{\tau}_i^0$, in group $i-1$ is $1/\gamma_{i-1}$, and so on until in group $I_{J(i)}$, where they only recall partners for the amount of time

$$t_{M_{J(i)}} = T_M - G_{J(i)+1,i} = (T_M - \bar{\tau}_i^0) - \sum_{l=J(i)+1}^{I-1} \frac{1}{\gamma_l}.$$

Hence

$$\alpha_i \approx \beta_{J(i)} t_{M_{J(i)}} + \sum_{k=J(i)+1}^{i-1} \frac{\beta_k}{\gamma_k} + \beta_i \bar{\tau}_i^0. \tag{3.11}$$

Case 3. $J(i) = 0$.

In this case, T_M is longer than the time the infected people have been infected. The identified infected individuals can identify all of the partners since they have been infected. As a result,

$$\alpha_i \approx \sum_{k=1}^{i-1} \frac{\beta_k}{\gamma_k} + \beta_i \bar{\tau}_i^0. \tag{3.12}$$

3.3. Estimation of O_i

For this first version of the model we neglect O_i in both models. The O_i people are those who are named as partners of infected people in the past T_M years, and became infected after that partnership. We estimate this term in Section 7, and show it is relatively small compared to L_i and M_i for the parameter ranges of interest. We justify neglecting this term by assuming that the rate of infection in the population is low enough, and the time period T_M is small enough, that the chances a person would contact an infected person and become infected after their contact with

the known infected is small compared to their chances of being infected by the infected they are known to have contacted. There may be situations outside of our parameter ranges where our assumptions are not valid, and these terms need to be included, but they are not examined in this paper.

4. The reproductive number and endemic equilibrium

We now summarize the results for both the reproductive number and the endemic equilibrium for the SCT-DI model and the SCT-SP model in this section. The details can be found in Appendices A and B. In the numerical results section we will use these results to examine the behavior and sensitivity of our two models.

4.1. The reproductive number

The reproductive number for the SCT-DI model is given by

$$R_0^{\mathrm{D}} = r\sum_{i=1}^{n} p_i\beta_i\tau_i^0, \tag{4.1}$$

where $\tau_i^0 = 1/(\mu + v_i + \sigma + f\sigma C_i^0)$.

Similarly, the reproductive number for the SCT-SP model has the same form as for the model without control measures:

$$R_0^{\mathrm{S}} = r\sum_{i=1}^{n} q_i\beta_i\tau_i^0, \tag{4.2}$$

where we define

$$q_i := \prod_{j=1}^{i-1}\gamma_j\tau_i^0, \tag{4.3}$$

$\tau_i^0 = 1/(\mu + \gamma_j + \sigma + f\sigma M_j^0)$, and M_i^0 is M_i evaluated at the infection-free equilibrium.

Note that in order to numerically determine the reproductive number for the SCT-SP model we need to first determine M_i^0. Recall that there are three different possible cases for these M_i. Hence we need to be careful when we evaluate them using the appropriate formula for the ith group. In Appendix A, we explicitly give R_0^S and q_i for some specific cases of M_i.

The partial derivatives of the reproductive numbers with respect to the rate of random screening, σ, and the fraction of identified partners contact traced, f, are given by

$$\begin{aligned}
\frac{\partial R_0^{\mathrm{D}}}{\partial\sigma} &= -r\sum_{i=1}^{n} p_i\beta_i(\tau_i^0)^2\left(1 + fr\widetilde{T}_i\beta_i\right),\\
\frac{\partial R_0^{\mathrm{S}}}{\partial\sigma} &= -r\sum_{i=1}^{n} q_i\beta_i\tau_i^0\left(\sum_{j=1}^{i}(1 + fM_j^0)\tau_j^0\right),
\end{aligned} \tag{4.4}$$

and

$$\frac{\partial R_0^{\mathrm{D}}}{\partial f} = -r\sum_{i=1}^{n} p_i\beta_i(\tau_i^0)^2\left(r\widetilde{T}_i\beta_i\sigma\right),$$
$$\frac{\partial R_0^{\mathrm{S}}}{\partial f} = -r\sum_{i=1}^{n} q_i\beta_i\tau_i^0\left(\sum_{j=1}^{i} M_j^0\sigma\tau_j^0\right). \tag{4.5}$$

All these derivatives are negative. Hence, both a pure random screening program (with $f = 0$) and any contact tracing program will reduce the reproductive number of the epidemic, and thus most likely reduce the severity of the epidemic. The more people are screened (the greater σ is), and the more partners people can recall or more accurate information people give (the greater f is), the more R_0 will be reduced for both models. A large enough screening rate and partner recall will reduce R_0 below the threshold.

Notice that contact tracing has a different impact on the reproductive number, and hence on the transmission dynamics, for the SCT-DI model than the SCT-SP model. It is clear from (4.4) and (4.5) that how contact tracing can reduce R_0 for the SCT-DI model. However, the contact tracing for the SCT-SP model depends on not only the time period that identified infected people can identify their partners back to but also how long they have been infected, which determines how many infected partners they have had.

4.2. The endemic equilibrium

For the SCT-DI Model, the endemic equilibrium is given by

$$S^* = \frac{\mu G(\hat{x})}{\mu G(\hat{x}) + F(\hat{x}) - 1}S^0, \tag{4.6}$$

$$I_i^* = \frac{\mu p_i(F(\hat{x}) - 1)}{(a_i + b_i\hat{x})(\mu G(\hat{x}) + F(\hat{x}) - 1)}S^0, \tag{4.7}$$

where

$$G(\hat{x}) := \sum_{i=1}^{n}\frac{p_i}{a_i + b_i\hat{x}}, \quad F(\hat{x}) := r\sum_{i=1}^{n}\frac{\beta_i p_i}{a_i + b_i\hat{x}},$$

and $\hat{x}$ is the (unique) root of the equation $H_{\mathrm{D}}(x) = 1$. Here $H_{\mathrm{D}}(x)$ is defined by

$$H_{\mathrm{D}}(x) = r\sum_{i=1}^{n}\frac{\beta_i p_i}{\frac{a_i}{x} + b_i}, \tag{4.8}$$

with

$$a_i = \mu + v_i + \sigma + rf\sigma T_{\mathrm{M}}, \quad b_i = f\sigma r(\widetilde{T}_i\beta_i - T_{\mathrm{M}}).$$

For the SCT-SP model, the endemic equilibrium is given by

$$I_n^* = \frac{\mu S^0}{\left(\mu\frac{\sum_{i=1}^{n}\prod_{j=i+1}^{n}\left(A_j + B_j\tilde{x}\right)}{1/\tilde{x} - 1} + \prod_{j=1}^{n}\left(A_j + B_j\tilde{x}\right)\right)}, \tag{4.9}$$

$$I_i^* = \prod_{j=i+1}^{n} \left(A_j + B_j\tilde{x}\right) I_n^*, \quad i = 1, \ldots, n-1, \tag{4.10}$$

$$S^* = \frac{\sum_{i=1}^{n} \prod_{j=i+1}^{n} \left(A_j + B_j\tilde{x}\right)}{1/\tilde{x} - 1} I_n^*, \tag{4.11}$$

where $\tilde{x}$ is the unique root of the algebraic equation

$$H_S(x) = rx \sum_{i=1}^{n} \frac{\beta_i}{\prod_{j=1}^{i}(A_j + B_j x)} - 1, \tag{4.12}$$

$$A_i = (\gamma_i + \mu + \sigma + f\sigma r T_M)/\gamma_{i-1}, \quad B_i = f\sigma r (J_{M_i} - T_M)/\gamma_{i-1},$$

with $\gamma_0 = 1$ and J_{M_i} determined from equations (3.10), (3.11) and (3.12). That is

$$J_{M_i} = \begin{cases} \beta_i T_M, & \text{if } J(i) = i \text{and } T_M \leqslant \bar{\tau}_i, \\ \beta_J(i) t_{M_{J(i)}} + \sum_{k=J(i)+1}^{i-1} \frac{\beta_k}{\gamma_k} + \beta_i \bar{\tau}_i, & \text{if } 1 \leqslant J(i) < i, \\ \sum_{k=1}^{i-1} \frac{\beta_k}{\gamma_k} + \beta_i \bar{\tau}_i, & \text{if } J(i) = 0. \end{cases} \tag{4.13}$$

The details can be found in Appendix B.

5. Numerical investigation of the models

Tables 2 and 3 give the parameter values we use for the original basic DI and SP models. We estimated these parameters in [8,9] from the published literature. Here we use the baseline parameters given in [9], which ensure that the two models have the same value of $\bar{\tau}$, and nearly identical values for R_0 and $\bar{\beta}$. Thus they have nearly identical endemic states, because the sensitivity of the models to the intervention programs can be better compared if these values are the same in the absence of any intervention program ($\sigma = 0$).

Because we are considering a high risk population, we assume that individuals realize they are at risk and are much more likely to come in for testing than in the general population, and that they are reasonably likely to change behaviors. We define the screening rate σ as the product of the fraction screened per year (ε) and the fraction who change behaviors (κ). In our simulations, we use a 5% average as a baseline screening rate per year ($\sigma = 0.05$), and study the sensitivity of the model to screening rates between 0 and 20%. We take $T_M = 1$ year. This value is consistent with many contact tracing programs, although six months is also a common look-back time span. We study the sensitivity of the model to variations in T_M between 0 and 2 years. In active populations, the fraction of partners named, located, and screened varies widely. Some programs seem to have no difficulty locating partners, but find many of them reluctant to be tested. Other programs have more difficulty locating partners, and less difficulty getting them to be tested [18]. For our simulations we assume that half of all partners from the past T_M years will be tested ($f = 0.5$), and study the sensitivity of the models to variations in f since some studies cited in [18] were more successful ($f > 0.5$), but others did worse. In none of these studies is there a way to evaluate the fraction of their partners that individuals actually named, because it would be very

Table 2
These parameters were chosen based on the studies and calculations cited in the text

Basic parameter	Formula	Value
Sexually active removal rate	α	0.05 yrs^{-1}
Natural death rate	d	0.02 yrs^{-1}
Mean duration of infection (when $\alpha = 0$ in the DI model)	$\bar{\tau}$	12 years
Partner acquisition rate	r	5 partners/yr
Contacts per partner parameter	η	1.0
Initial population size	$N(0)$	S^0
Initial infected population	$I_T(0)$	0.01 S^0
Normalized infection-free equilibrium	S^0	1
DI parameters		
Distribution of the newly infected	p	(0.05, 0.33, 0.5, 0.12)
Progression rates by group	v	(0.19, 0.096, 0.058, 0.028) yrs^{-1}
Relative per contact transmission	ζ	$(10^3, 10^2, 10, 1)\, z^{\mathrm{D}}$
Infectivity adjustment factor	z^{D}	5.1×10^{-5}
SP parameters		
Progression rates by group	γ	(13.0, 0.23553, 0.23553, 0.47) yrs^{-1}
Relative per contact transmission	ζ	$(100, 1, 1, 10)\, z^{S}$
Infectivity adjustment factor	z^{S}	9.08×10^{-4}

Note that these models allow the population to be normalized such that $S^0 = 1$.

Table 3
Derived parameters

Description	Formula	Baseline value
Duration of infection	$\bar{\tau}$	7.3 yrs
Mean probability of transmission per contact	$\bar{\zeta}$	0.003
Number of contacts per partner	$c(r = 5)$	21.8 contacts per partner
DI parameters		
Probability of transmission per partner	β	(0.68, 0.105, 0.011, 0.0011)
Mean probability of transmission per partner	$\bar{\beta}$	0.053
Reproductive number	R_0	1.93
SP parameters		
Probability of transmission per contact	β	(0.87, 0.0196, 0.0196, 0.1802)
Mean probability of transmission per contact	$\bar{\beta}$	0.051
Reproductive number	R_0	1.88

These parameters are derived from the parameters given in Table 2.

difficult to determine how good high risk people's memories are when it comes to recalling their sexual partners, or how often they deliberately leave someone off their list.

Estimates of the mean probability of infection per contact, $\bar{\zeta}$, range from 0.0003 (lowest value estimated for female-to-male transmission) to 0.08 (highest value estimated for male-to-male transmission) [27]. Here we use $\bar{\zeta} = 0.003$ at baseline.

In this section, we investigate the effectiveness of these simple random screening and contact tracing programs for three levels of interventions: none, random screening only, and random screening plus contact tracing. Next, we use the analytical formulas for R_0 to analyze the sensitivity of the early epidemic to different levels of intervention programs by varying σ, f, and T_M. We then examine the sensitivity of the long-term epidemic to these three parameters. Finally, we investigate the impact of our approximations for the SCT-SP contact tracing model on the smoothness of R_0 and the endemic equilibrium.

The impact of these interventions on the SCT-DI and SCT-SP epidemics shows how the effectiveness of the intervention strategy depends on the underlying etiology of the disease transmission. These simulations confirm that contact tracing is more effective when there are core groups which are transmitting the majority of the infections (as in the SCT-DI model) than when a large fraction of the infections are spread by those who have just been infected (as in the SCT-SP model). Contact tracing is less effective in the SCT-SP model, because the largest fraction of infections in our simulations is caused by those who have been infected the longest, and contact tracing may be too late in identifying these individuals. It is also interesting that, while contact tracing would appear to be an effective approach to reducing the overall spread of infection in the SCT-DI model, the relative importance of the most infectious group to the spread of the infection remains the same as it was without the contact tracing. We conclude that if the impact of the intervention program depends on the underlying etiology of the infection, this etiology must be understood in order to design the cost-effective intervention programs.

The timing of a multigroup model epidemic is extremely sensitive to the initial distribution of the infected population. The initial conditions should also be consistent with the assumptions used to define the quantity $M_i + L_i$ in the contact tracing model. We defined the initial distribution of the 1% infected population using the *Numerical Preinitialization Procedure* described in [9]. This distribution is defined to simulate the behavior of a naturally occurring epidemic, and to minimize the initial transients created by artificial initial conditions. First, a tiny fraction (0.01%) of the population is distributed among the infected groups based on the relative fraction of time when an individual is in a particular group. That is, the I_i is initialized with 0.0001 $S^0\bar{\tau}_i/\bar{\tau}$, where $\bar{\tau}_i$ is the duration of infection of infected individuals in group i. The model is then advanced forward in time until 1% of the population has become infected. At that time, the total population is renormalized to equal S^0 and the time is renormalized for this point to be $t = 0$. The $I_i(0)$ are given the same relative distribution as they had when the simulation is stopped, and their sum is set to 0.01 S^0. This approach is an approximation of the natural initial conditions that would occur if a very small number of infected people were initially introduced into the population. It also sets up initial conditions which are consistent with the contact tracing terms in the models.

5.1. Transient dynamics of the models

The impact of random screening and contact tracing on the transient dynamics can be seen in Fig. 5. In the first simulation (solid lines), there is no intervention, and all parameters are at the

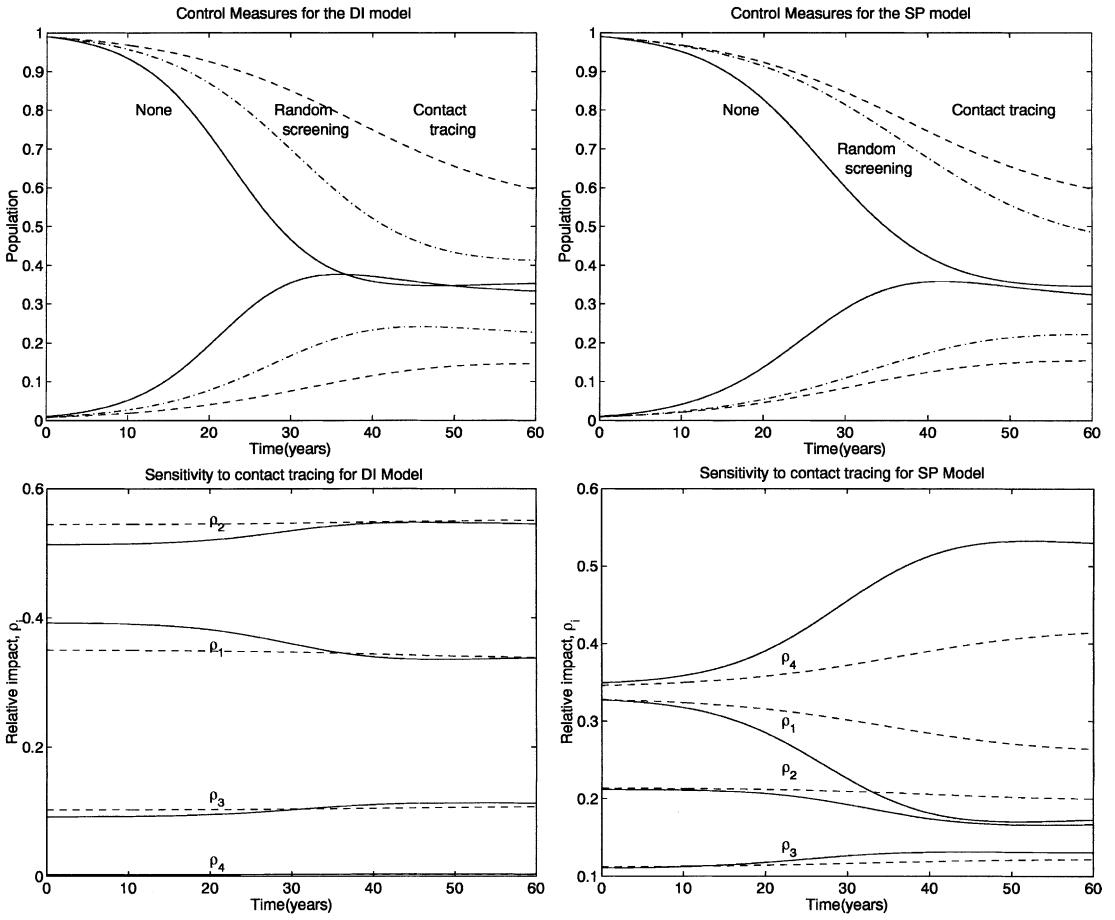

Fig. 5. The solid lines plot the susceptible and infected populations when there is no screening or contact tracing, the dash-dot lines are for the epidemics when 5% of the population is screened, and there is no contact tracing, and the dashed lines show what happens when contact tracing is added to the the random screening model with $T_M = 1$ year and $f = 0.5$. The upper figures show the susceptible and infected populations for these three cases for each model. In the original model runs, the DI epidemic is larger than the SP epidemic, despite having similar R_0 and endemic states. In the upper left figure, random screening has a modest impact in slowing the SCT-DI model epidemic when compared to the more dramatic impact of contact tracing. In the upper right figure, random screening alone has more impact on the SCT-SP model than the SCT-DI model. However, contact tracing has less impact on the SCT-SP model epidemic, so that the combined programs are about equally effective for the two models. The lower two figures show the relative impact, ρ_i, (see Eq. (2.6)), for each of the two models, for the baseline and contact tracing cases. Contact tracing changes the relative importance of the most infectious groups more in the SCT-SP model than in the SCT-DI model.

baseline values in Table 2. In the second simulation (dash–dot lines), there is screening of 5% of the active population and no contact-tracing. In the third simulation (dash lines), contact tracing is added to the 5% screening program, with $T_M = 1$ and $f = 0.5$.

In the SCT-DI model there is a significant impact from screening alone. Furthermore, a modest amount of contact tracing added to this screening program leads to another large reduction in the epidemic. The lower plots show the relative impact, ρ_i, defined as the fraction of infections caused

by group i. Surprisingly, in the SCT-DI model contact tracing has only a slight shift in the relative impact of the different groups on spreading the epidemic even though there is a huge reduction in the infected population.

The 5% random screening program has slightly more impact on the SCT-SP model than on the SCT-DI model. The relative impact plots illustrate that the contact tracing changes the underlying dynamics of the SCT-SP epidemic. With no intervention program, one third of infections early in the epidemic are caused by group 1, and most of infections late in the epidemic by group 4. Contact tracing identifies people before they enter group 4 and therefore group 1 has more relative impact on the epidemic throughout the epidemic than it does with no control program.

The group causing the most infections can impact which control methods will work best. Because in the SCT-SP model people stay in group 1 for such a short time, they are hard to detect. However, by the time they reach group 4, there is a reasonable chance that they know about their infection. This implies that contact tracing used in conjunction with an early identification program, such as a concerted effort to screen people who have early symptoms of infection, may be an effective intervention program for an SCT-SP epidemic.

5.2. Sensitivity of R_0

In Section 4, we determined that R_0 decreases for both models as either σ or f increases. Thus the more screening or contact tracing there is, the more slowly the initial epidemic will grow. To measure the sensitivity of the initial epidemic to the intervention programs, we evaluate R_0 using the baseline parameters given in Table 2, and varying the random screening rate, σ, the fraction of partners traced, f, and the time window for remembering past partners, T_M. These results are shown in Fig. 6. The upper figures show R_0 as a function of σ for five values of T_M, $(0, 0.5, 1, 1.5, 2)$, and $f = 0.5$. The lower figures show R_0 as a function of f for 5 values of σ, (0.0, 0.05, 0.1, 0.15, 0.2), and $T_M = 1$ year.

These figures show that R_0 is more sensitive to changes in σ than to variations in either f or T_M over their range. As random screening increases, R_0 for the SCT-DI model decreases more rapidly than R_0 for the SCT-SP model for the same level of contact tracing. The upper plots show that the SCT-SP model is less sensitive to T_M than the SCT-DI model. For the SCT-DI model at a 10% screening rate, R_0 drops as T_M increases crossing threshold conditions ($R_0 = 1$) before $T_M = 1.5$ years. On the other hand, if identified infected people can identify their partners for just one year, and half of their partners can be traced, then the SCT-DI model goes below threshold when about 12.5% of at risk people are randomly tested.

The lower graphs show that R_0 is less sensitive to the fraction of partners traced than to the random screening rate, and is more sensitive in the SCT-DI model than in the SCT-SP model. Note that, for the SCT-DI model, R_0 drops below threshold on the 10% random screening curve at f about 0.7, that is, if 70% of the past partners are traced and 10% population randomly screened, the epidemic is below threshold for the SCT-DI model, while R_0 never drops below 1 on the 10% random screening curve for the SCT-SP model.

Finally, we remark that additional studies have shown that in the SP model, R_0 remains in the range [1.3,1.5] for $\sigma = 0.05$, $f \in (0, 1)$, and $T_M \in (0, 2)$. In the SCT-DI model, R_0 decreases more rapidly, falling quickly at small values of f and T_M, and drops below threshold at larger values of f and T_M.

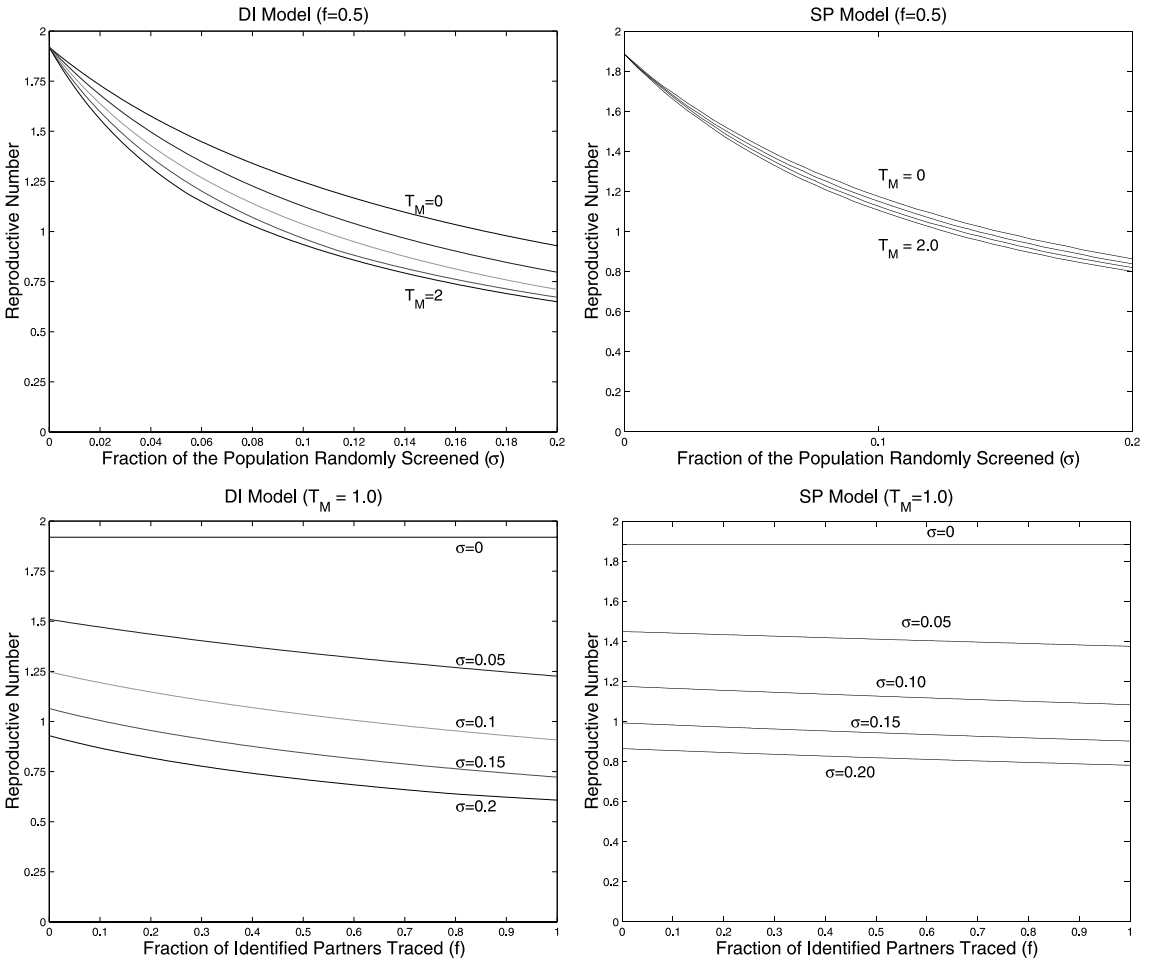

Fig. 6. In the top figures, R_0 is plotted as a function of the fraction of the population that is randomly screened for infection. The different curves illustrate how much greater the impact of contact tracing ($f = 0.5$) is for the DI model than the SCT-SP model for $T_M = 0$, 0.5, 1, 1.5, 2. To illustrate the sensitivity of the models to f in the lower figures, we fix $T_M = 1$ year and plot R_0 as a function of the fraction of partners traced, f. The multiple curves illustrate the impact when the fraction of the population randomly screened is varied, $\sigma = 0$, 0.05, 0.1, 0.15, 0.2. R_0 is reduced more in the SCT-DI model than in the SCT-SP model as the fraction of partners traced increases.

5.3. Sensitivity of the endemic equilibrium

We show in Appendix B that when $R_0 > 1$ there exists a unique endemic equilibrium for both models. We solve for the endemic equilibrium by numerically finding the roots of the algebraic equilibrium equations defined in Section 4. This is easily accomplished because x in (4.8) and (4.12) is an increasing function of σ, f, and T_M. The endemic equilibrium I_i^* given in (4.7), (4.9) or (4.10), is a function of σ, f, and T_M. Changes in these parameters affect I_i^* as functions of x and through the values of a_i and b_i for the SCT-DI model, or the values of A_i and B_i for the SCT-SP

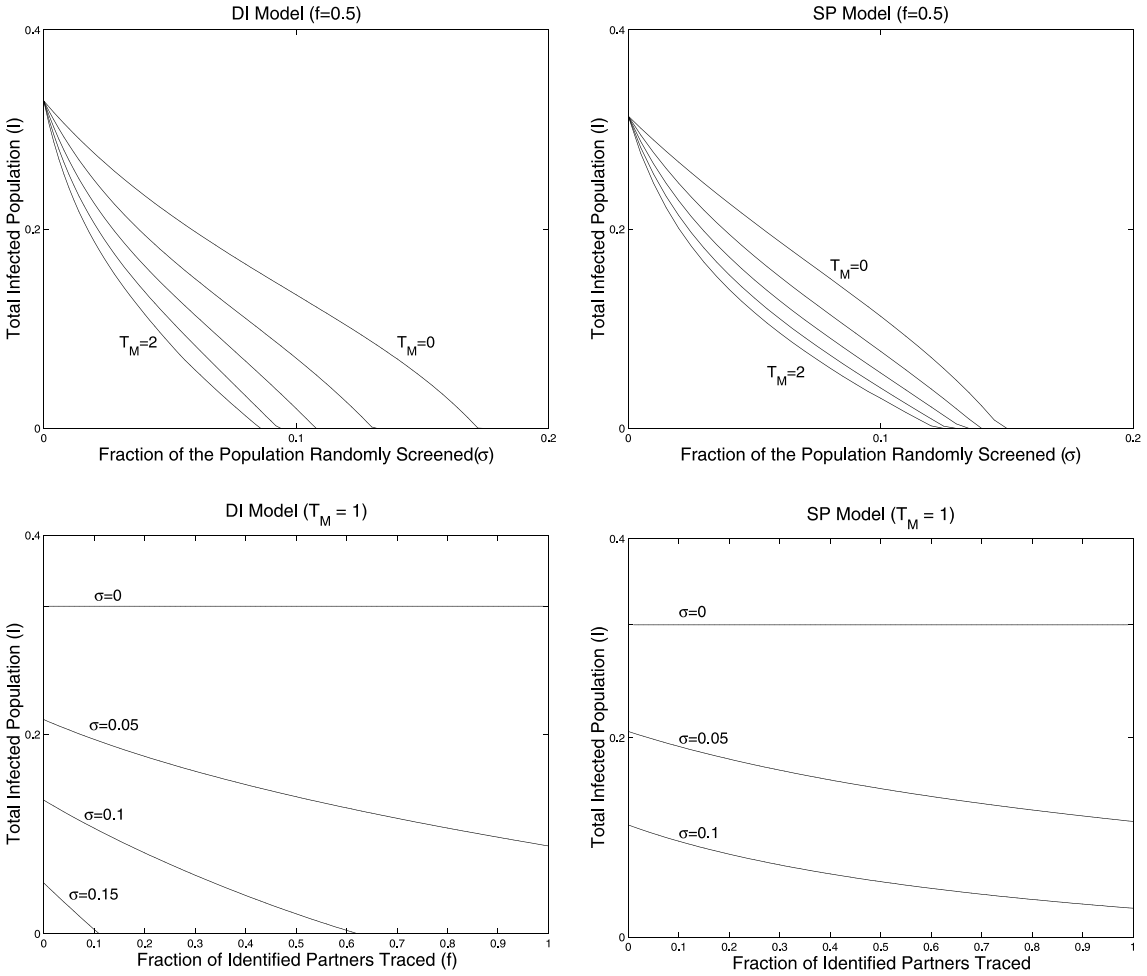

Fig. 7. To examine the sensitivity of the endemic equilibria, we plot the total infected population of the SCT-DI and SCT-SP models at the endemic equilibrium as we vary the fraction of the population randomly screened (σ) for 5 values of T_M and $f = 0.5$, and the fraction of the population traced (f) for 3–4 values of σ and $T_M = 1$ year. Notice that in all of these sensitivity studies, contact tracing has more impact on the endemic equilibrium for the SCT-DI model than for the SCT-SP model.

model. Because of these complex interrelationships, we investigate the sensitivity of the endemic equilibrium numerically and illustrate our results in Fig. 7.

In Fig. 7 we see that I^* is a decreasing function of all three parameters in the models. Whenever R_0 crosses the threshold values $R_0 = 1$, the total number of infected people at the endemic equilibrium vanishes and the lines on the graph intersect the x-axis. As in our studies of R_0, we find that the contact tracing program has more impact on the SCT-DI model epidemic than the SCT-SP model epidemic. For example, when $f = 0.5$ there is a more rapid decrease of I^* in the SCT-DI model than in the SCT-SP model. If there is no contact tracing, ($T_M = 0$), then screening alone has a bigger impact at slowing the epidemic in the SP model than in the DI model. As T_M is increased, the critical value of σ for stopping the epidemic decreases almost twice as fast for the SCT-DI model as for the

SCT-SP model. There is a similar response to increasing f and σ, at fixed T_M. If the screening rate is small and the SCT-DI model holds, a good contact tracing program can bring the epidemic under control.

5.4. Impact of the discrete approximations in the SCT-SP model

In developing the contact tracing SCT-SP model, we estimated how far back people can identify their partners. We assumed that the mean time an individual has been in a group is approximated by the mean time a typical individual stays in a group, $\bar{\tau}_i^0$. We also assumed that we can use the mean time that people stayed in previous groups to estimate how many past groups a person in group i can recall their partners from. The first of these approximations ignores variability in population sizes over time and is exact when the population is at equilibrium. The second assumption about how to compute averages leads to a possible discontinuity in the SCT-SP model as the parameter T_M changes and the index $J(i)$ jumps.

In Fig. 8, we investigate the nature of these jumps and show that they lead to small kinks, but not discontinuities, in R_0 and the endemic equilibrium. Both plots exhibit rapid drops at some T_M between 1.5 and 3 years, but this shift is short-lived, due to the discontinuous change in slope. Also the kinks occur after the time where most programs stop contact partners, ($T_M < 1$ year).

6. Validity of the model assumptions

To gain insight into the impact of contact tracing in reducing the spread of HIV, and develop a differential equation model which captures the main effects of such a program, we made a number of approximations in Section 3. We justified most of these approximations based on the time scales involved, and the processes of sexual transmission and contact tracing. We now examine whether our assumptions are valid, and describe how to extend the model to account for a more complex contact tracing process. Here we develop an expression for one term we neglected, O_i, and briefly discuss how the approximations to the time spent in each group could be improved upon. We show that when T_M is small, O_i is relatively small compared to the terms L_i and M_i which we included in the original model. We also discuss how to replace the approximation for $T_i(t)$ used in Section 3 with the true expression for $T_i(t)$, and some of the difficulties which would be encountered in doing so.

6.1. Estimation of O_i

In the SCT models in Section 3, we assumed that O_i was negligible compared to L_i and M_i. O_i is the average number of partners named by a screened infected who became infected between the time they had contact with that person and the time that person was screened. The combined events of not becoming infected by their contact with one person, and then subsequently having contact with other infected people and becoming infected by one of them, would be, in general, fairly rare events. However, in our simulations we are dealing with a population with 5 new partners per year, and a high level of infections. It is possible that under different assumptions the term O_i may have a significant impact on the model behavior.

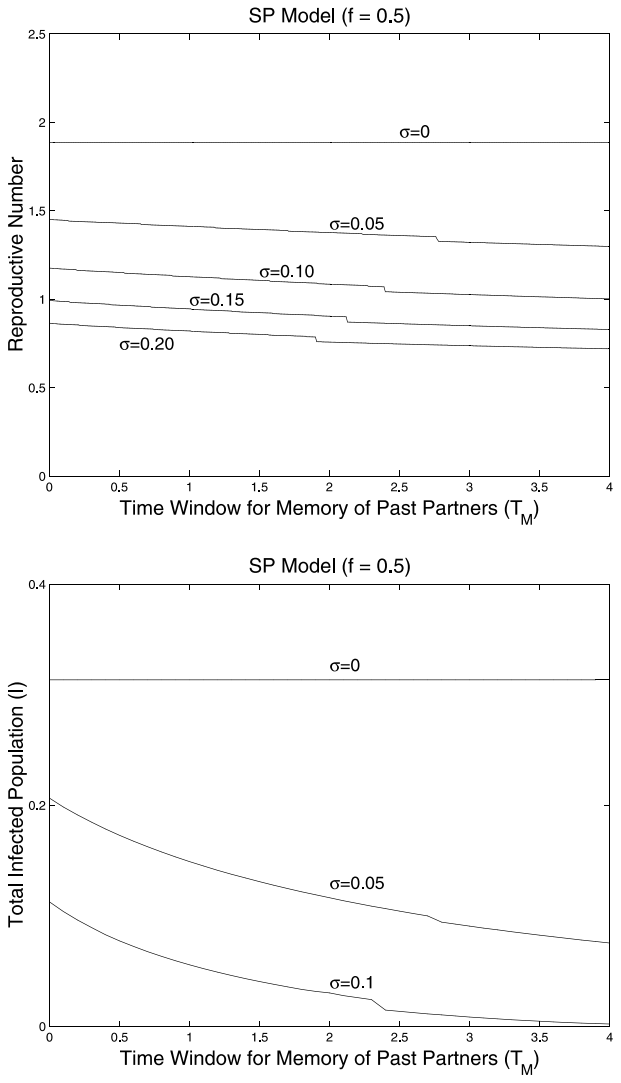

Fig. 8. To examine the impact of the discontinuity in the parameter T_M that is introduced into the SCT-SP model by the model approximations and the structure of the SCT-SP model, we plot R_0 and the total number of infecteds at equilibrium as a function of T_M for different values of σ. We see that in fact these quantities are continuous in T_M. There is a discontinuity in their slopes for T_M between 1.5 and 3 years, but the drops are small, and the curves quickly return to their former behavior.

6.1.1. $O_i(t)$ for the SCT-DI model

Recall that $T_i(t)$ is the mean time that an infected person in group I_i has been in that group, and $\widetilde{T}_i(t) = \min\{T_M, T_i(t)\}$ is the minimum of the time the person is asked about and the time he has been infected. The average number of susceptible people a screened infected person from the group had contact with in the past $\widetilde{T}_i(t)$ years, but did not infect, is then

$$r(1-\beta_i)\int_{\widetilde{t-T_i(t)}}^{t}\frac{S(s)}{N(s)}\,\mathrm{d}s.$$

Since the contact with a susceptible occurred at time s in this integral, after time s the probability that the susceptible becomes infected before time t is

$$r\sum_{j=1}^{n}\beta_j\int_s^t\frac{I_j(\tau)}{N(\tau)}\,\mathrm{d}\tau.$$

Thus

$$O_i(t)=r(1-\beta_i)\int_{\widetilde{t-T_i(t)}}^{t}\frac{S(s)}{N(s)}r\sum_{j=1}^{n}\beta_j\int_s^t\frac{I_j(\tau)}{N(\tau)}\,\mathrm{d}\tau\,\mathrm{d}s.$$

Just as in the original SCT models, we make the simplifying assumption that S and N can be approximated by their time t values for the length of time $\widetilde{T}_i(t)$, giving

$$O_i(t)\approx r^2(1-\beta_i)\frac{\widetilde{T}_i^2S(t)}{2N^2(t)}\sum_{j=1}^{n}\beta_jI_j(t). \tag{6.1}$$

Splitting this into two factors, as we did with M_i, we have

$$O_i(t)\approx\frac{1}{2}\lambda(t)\frac{rS(t)}{N(t)}\phi_i \tag{6.2}$$

where $\phi_i=(1-\beta_i)\widetilde{T}_i^2$.

6.1.2. $O_i(t)$ for the SCT-SP model

The expression for $O_i(t)$ in the SCT-SP model is very similar to the integral expression given above for the SCT-DI model. The only difference is that, since the screened infecteds have potentially gone through previous stages of the disease, their infectivity may have changed, and thus β_i becomes a function of s and must be integrated. This gives

$$O_i(t)\approx r\int_{\widetilde{t-T_i(t)}}^{t}(1-\beta(s))\frac{S(s)}{N(s)}r\sum_{j=1}^{n}\beta_j\int_s^t\frac{I_j(\tau)}{N(\tau)}\,\mathrm{d}\tau\,\mathrm{d}s.$$

Again, making the approximations that the populations have reached their t values over the course of this integral, we have

$$O_i(t)\approx\frac{1}{2}\lambda(t)\frac{rS(t)}{N(t)}\phi_i,$$

where

$$\phi_i=\int_{\widetilde{t-T_i(t)}}^{t}(1-\beta(s))\int_s^t\mathrm{d}\tau.$$

We estimate ϕ_i for different groups as follows:

For group 1, since screened infecteds from group I_1 have only entered one group after becoming infected, the estimation of ϕ_1 is the same procedure as for the DI model, giving

$$\phi_1(t)\approx(1-\beta_1)\widetilde{T}_1^2.$$

For the remaining groups ($i > 1$), we again consider the three possible cases, and do a similar set of approximations as we did for α_i in Section 3.2.

Case 1. $J(i) = i$ and $T_M \leqslant \bar{\tau}_i$.

In this case, the average infected person arrived in their current infected group so long ago that they cannot identify partners they had while they were in a previous group. Then

$$\phi_i \approx (1-\beta_i)T_M^2.$$

Case 2. $1 \leqslant J(i) < i$.

In this case, T_M is longer than the time people have been in group I_i, but shorter than the average time they have been infected. Let $\beta(u)$ be the infectiousness that the 'index case', who is now in group I_i, used to have at the previous time u. Then, since we trace back in the time interval the index case can recall, $\beta(u)$ goes backwards with $\beta(0) = \beta_i$ and $\beta(T_M) = \beta_{J(i)}$. Note that this is a further approximation. To do this exactly we would have to reframe the infected population in terms of a time since infection. However, this estimate is adequate for our purposes here. Then

$$\begin{aligned}\phi_i &= 2\int_{t-T_M}^{t}(1-\beta(t-s))\int_s^t \mathrm{d}\tau\,\mathrm{d}s \\ &= (1-\beta_i)\bar{\tau}_i^2 + \sum_{k=J(i)+1}^{i-1}(1-\beta_k)\left(\left(\bar{\tau}_i + \sum_{l=k}^{i-1}\frac{1}{\gamma_l}\right)^2 - \left(\bar{\tau}_i + \sum_{l=k+1}^{i-1}\frac{1}{\gamma_l}\right)^2\right) \\ &\quad + \left(1-\beta_{J(i)}\right)\left(\left(\bar{\tau}_i + \sum_{l=J(i)+1}^{i-1}\frac{1}{\gamma_l} + t_{M_{J(i)}}\right)^2 - \left(\bar{\tau}_i + \sum_{l=J(i)+1}^{i-1}\frac{1}{\gamma_l}\right)^2\right)\end{aligned}$$

Using the $G_{i,k}$ defined in Section 3.2, we can simplify this expression as

$$\phi_i = (1-\beta_i)\bar{\tau}_i^2 + \sum_{k=J(i)+1}^{i-1}(1-\beta_k)(G_{k,i}^2 - G_{k+1,i}^2) + \left(1-\beta_{J(i)}\right)(T_M^2 - G_{J(i)+1,i}^2).$$

Case 3. $J(i) = 0$.

In this case, T_M is longer than the time the infected people have been infected. The identified infecteds can identify all of their partners since they have been infected. As a result

$$\phi_i = (1-\beta_i)\bar{\tau}_i^2 + \sum_{k=1}^{i-1}(1-\beta_k)\left(\left(\bar{\tau}_i + \sum_{l=k}^{i-1}\frac{1}{\gamma_l}\right)^2 - \left(\bar{\tau}_i + \sum_{l=k+1}^{i-1}\frac{1}{\gamma_l}\right)^2\right),$$

or

$$\phi_i = (1-\beta_i)\bar{\tau}_i^2 + \sum_{k=1}^{i-1}(1-\beta_k)(G_{k,i}^2 - G_{k+1,i}^2).$$

Notice that all O_i are positive terms. They would increase the impact of contact tracing. However, as shown in Fig. 9 including these terms has only a small effect. So long as T_M is within the realm of most programs, i.e. one year or less, the most we may possibly lose would be about 5% of the effect of the total program. Thus, it is reasonable to neglect them in our model formulation unless we wish to study a population with a longer T_M.

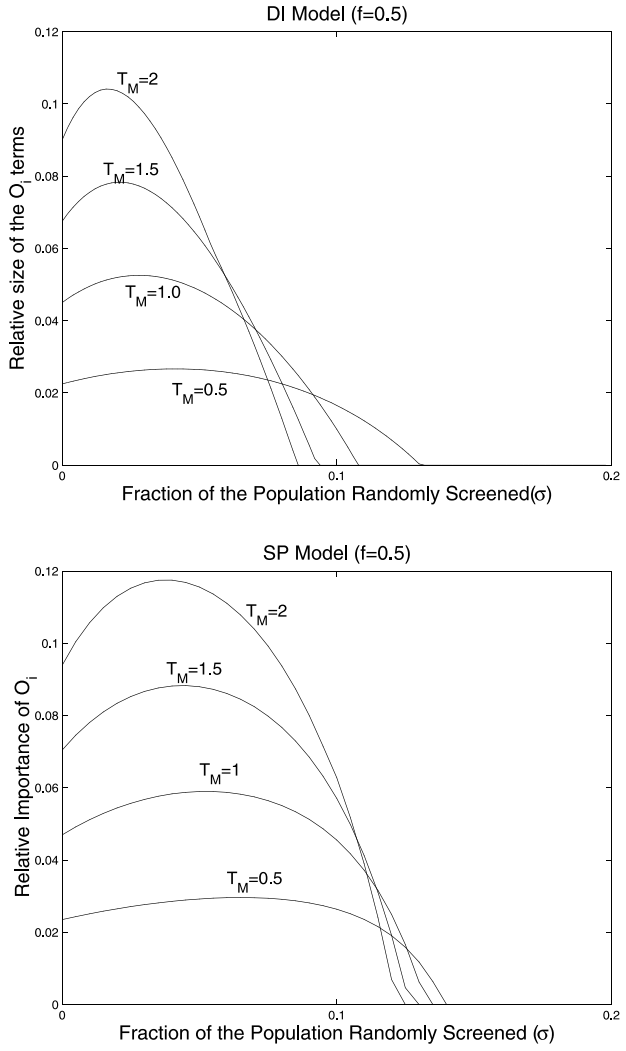

Fig. 9. This shows $\sum_{i=1}^{4} O_i / \sum_{i=1}^{4} C_i$ as computed at the equilibrium values from the models without O_i (i.e. the models used in Section 3). We see that for $T_M = 2$ years, this quantity is larger than 0.1 in the worst case for both models, showing that for extremely large $T_M > 2$, the actual impact of contact tracing may be as much as 10% greater than our model predicts. However, for more realistic values of T_M (1 year or less), these terms remain relatively small, and we are justified in neglecting them.

Note also that all O_i include a factor of λ, and thus are zero at the infection-free equilibrium. This implies that they do not affect the reproductive number, although they do shift the endemic equilibrium values.

6.2. Estimation of $T_i(t)$

The second approximation we examine is the one we made for the mean time that an infected person has been in their current group, $T_i(t)$. This was estimated by $\bar{\tau}_i^0 = (\mu + v_i + \sigma + f\sigma C_i^0)^{-1}$ for

the SCT-DI model, and by the same formula with v_i replaced by γ_i for the SCT-SP model, where it appears as τ_i^0 throughout our formulas. To accurately determine these quantities would require adding another variable to our model: the time since the person entered their current group. This would then turn the differential equations for the infected populations into partial differential equations. If τ is the new variable indicating the duration of time since the person entered infected group i, the infecteds become functions of two variables, t and τ, such as $I_i = I_i(t,\tau)$. Then,

$$T_i(t) = \int_0^\infty \tau I_i(t,\tau)\,\mathrm{d}\tau \left(\int_0^\infty I_i(t,\tau)\,\mathrm{d}\tau \right)^{-1}.$$

Because the rate of leaving the infected populations contains the term $C_i(t)$, the full model leads to an integro-partial-differential equation model, and its mathematical analysis becomes much more complicated. Such a model is beyond the scope of this paper.

However, because the estimate we have chosen at the infection-free equilibrium is valid, the estimate at the endemic equilibrium for T_i would have the same formula, with $C_i(0)$ replaced by its value at the endemic equilibrium. Numerical studies have shown that replacing $C_i(0)$ in the formula for $T_i(t)$ by $C_i(t)$ only affects results by less that 1%. Nevertheless, the case where using the actual formula for T_i could potentially have a large effect is in the early epidemic when people have only been infected a short period of time. Thus, in this early epidemic, T_i is much smaller than τ_i^0, is increasing rapidly, and will depend greatly upon assumptions about the initial distribution of infections in τ.

7. Summary and conclusions

We have investigated how mathematical models can help predict the effectiveness of control measures on the spread of HIV and other STDs. We studied the impact of random screening and contact tracing within the context of two HIV transmission models. In the DI model the infected population is divided into groups according to their infectiousness, and HIV is primarily spread by a small highly infectious group of superspreaders. Random screening alone reduces the impact of the epidemic a small amount for this model, while contact tracing slows the epidemic significantly by identifying the superspreaders. In the SP model an infected individual goes through a series of infection stages, and the virus is primarily spread by individuals in an initial highly infectious stage or in the late stages of the infection. In the SP model we find that adding contact tracing to random screening causes only a small decrease in infections compared to the decrease obtained by adding random screening to the uncontrolled epidemic. This occurs because contact tracing cannot identify very many of the people in the very short, initial, most infectious period. Thus the effectiveness of the intervention strategy strongly depends on the underlying etiology of the disease transmission.

While the terms that account for random screening are easily included in models of disease transmission, it is not obvious how to account for contact tracing. At first glance it would appear that, because contact tracing involves identifying events which occurred in the past, a model that includes it would contain nested integrals over the past. These integrals would be analytically intractable. However, by making a few simplifying but reasonable assumptions, we showed that these integrals could be approximated by functions which do not depend upon the past. This

allowed us to derive differential equation models directly from the epidemiology of the disease. These models have the advantage that we are able to determine analytical formulas for the reproductive numbers and endemic equilibria, and use those formulas to quickly gain insight into how effective contact tracing would be as part of an intervention program.

Using our results on the reproductive number and endemic equilibrium, we analyzed the impact of various levels of intervention programs on the early epidemic and the endemic equilibrium. We simulated the time evolution of several scenarios, and examined the effectiveness of contact tracing in identifying the most infectious group transmitting the infection. These studies led us to the following conclusions:

- Random screening and contact tracing can be included in simple STD differential equation transmission models.
- Contact tracing is most effective when there are core groups of individuals, remaining in the high risk population for long periods of time, that are transmitting the majority of the infections (as in the DI model).
- Random screening plus contact tracing is only slightly more effective than random screening alone when a large fraction of the infections are transmitted by individuals in a short, highly infectious, early stage within the disease progression (as in the SP model).
- When using models to guide intervention strategies, the underlying eitiology of the disease transmission must be captured by the model before it can be used to estimate the impact of the intervention on the epidemic.

The insights gained by any model are only as valid as the quality of the approximations and assumptions which go into developing the model. This is particularly true when modeling a complex situation, such as contact tracing. In order to investigate the validity of our model, we have, in Section 6, obtained expressions for the two most important terms that we neglected in the model development, and shown that, under realistic parameter estimates, one of these terms is indeed small compared to the one we kept.

We have described how mathematical models based on the transmission mechanisms of HIV can help the scientific community evaluate the potential effectiveness of different approaches for bringing an epidemic under control. It would be possible for public health officials or economists to add dollar amounts to various levels of screening and contact tracing in a particular population, and estimate the cost of reducing the epidemic to certain levels using these two models.

However, we caution that one must be careful about making policy recommendations based upon a single model until results have been validated via other modeling approaches or field studies, especially for something as important and politically volatile as methods for controlling the spread of HIV. The models formulation and analysis in this paper are only a preliminary study in this direction. Before any conclusions can be applied to a real-world setting, the model should account for variations in sexual behavior and in mixing patterns to our model. Results should also be directly compared with those from individual agent-based models. Individual agent based models for contact-tracing are based on different assumptions than differential-equation models, and may lead to different conclusions. Although it is difficult to derive analytical expressions for threshold conditions and equilibria for agent-based models, and the models require many simulations in order to obtain the distribution for each set of parameter values, the agent-based models

have the advantage that they can account exactly for the history of contacts between any two individuals. Therefore, it is important to study the impact of contact tracing using a variety of techniques, and this may lead to a more complete understanding of the spread of HIV and the effects of random screening and contact tracing on that spread under different conditions.

Although we have separated the DI and SP mechanisms in order to understand each of their roles, it appears from the data that HIV infected people both go through stages and have different individual levels of virus during the chronic infection stage. The real model should be a combined DI and SP model, which we will study in a future paper. Thus these insights are just one step in improving our understanding of the essential relationships between the social and biological mechanisms that influence the spread of the disease and can help set priorities in research, saving time, resources, and lives.

Appendix A. The reproductive number

We define the reproductive number R_0 such that the infection-free equilibrium is asymptotically stable if $R_0 < 1$ and is unstable if $R_0 > 1$.

A.1. R_0 for the SCT-DI model

The Jacobian of the contact tracing DI model (3.1) at the infection-free equilibrium can be written in the form

$$\begin{pmatrix} -\mu & \cdot & 0 \\ 0 & J_{DI} & 0 \\ 0 & \cdot & D_1 \end{pmatrix}, \tag{A.1}$$

where

$$D_1 = \operatorname{diag}(-(\mu + v_1), \ldots, -(\mu + v_n)),$$

and

$$J_{\mathrm{DI}} = \begin{pmatrix} p_1 r\beta_1 - \delta_1 & p_1 r\beta_2 & \cdots & p_1 r\beta_n \\ p_2 r\beta_1 & p_2 r\beta_2 - \delta_2 & \cdots & p_2 r\beta_n \\ \vdots & \vdots & \ddots & \vdots \\ p_n r\beta_1 & p_n r\beta_2 & \cdots & p_n r\beta_n - \delta_n \end{pmatrix}, \tag{A.2}$$

with $\delta_i = \mu + v_i + \sigma + f\sigma M_i^0 = 1/\tau_i^0$. Here $M_i^0 = r\widetilde{T}_i\beta_i$ is M_i in (3.7) evaluated at the infection-free equilibrium. Because all of the entries of the diagonal submatrix D_1 are negative, the stability of (A.1) is determined by J_{DI}.

Using the same approach as in [8] to analyze the matrix J_{DI}, it is a straightforward calculation to obtain the reproductive number

$$R_0^{\mathrm{D}} = r\sum_{i=1}^{n} \frac{p_i\beta_i}{\delta_i} = r\sum_{i=1}^{n} p_i\beta_i\tau_i^0 \tag{A.3}$$

for the DI model.

A.2. R_0 for the SCT-SP model

The Jacobian at the infection-free equilibrium for the contact tracing SP model (3.2) can also be written in the form

$$\begin{pmatrix} -\mu & \cdot & 0 \\ 0 & J_{SP} & 0 \\ 0 & \cdot & D_2 \end{pmatrix}, \tag{A.4}$$

where

$$D_2 = \operatorname{diag}(-(\mu+\gamma_1),\ldots,-(\mu+\gamma_n)),$$

and

$$J_{\mathrm{SP}} = \begin{pmatrix} r\beta_1-\delta_1 & r\beta_2 & r\beta_3 & \cdots & r\beta_{n-1} & r\beta_n \\ \gamma_1 & -\delta_2 & 0 & \cdots & 0 & 0 \\ 0 & \gamma_2 & -\delta_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \gamma_{n-1} & -\delta_n \end{pmatrix}, \tag{A.5}$$

with $\delta_i = \mu+\gamma_i+\sigma+f\sigma M_i^0 = 1/\tau_i^0$ and M_i^0 being M_i for the SP model evaluated at the infection-free equilibrium. Again, the stability of (A.4) is determined by that of matrix J_{SP} in (A.5).

Using a similar approach as in the derivation of R_0 for the SP model in [8] we can express the reproductive number for (3.2) as

$$R_0^{\mathrm{S}} = r\sum_{i=1}^{n}\frac{\beta_i q_i}{\delta_i} = r\sum_{i=1}^{n} q_i\beta_i\tau_i^0, \tag{A.6}$$

where

$$q_i := \prod_{j=1}^{i-1}\gamma_j\tau_i^0. \tag{A.7}$$

Note that M_i^0 depends on how far back people can identify their partners. If T_{M} is small and a typical infected person can only identify partners from their current infection stage ($J(i)=i$) and $T_{\mathrm{M}} < 1/(\mu+\gamma_i)$, then $M_i^0 = rT_{\mathrm{M}}\beta_i$ and

$$R_0^{\mathrm{S}} = r\sum_{i=1}^{n}\frac{q_i\beta_i}{\mu+\gamma_i+\sigma+f\sigma rT_{\mathrm{M}}\beta_i},$$

where

$$q_i := \prod_{j=1}^{i-1}\frac{\gamma_j}{\mu+\gamma_j+\sigma+f\sigma rT_{\mathrm{M}}\beta_j}.$$

At the other extreme if T_{M} is large and a typical infected person can identify all their partners, then

$$M_i^0 = r\left(\sum_{k=1}^{i-1}\frac{\beta_k}{\gamma_k}+\frac{\beta_i}{\mu+\gamma_i}\right),$$

and

$$R_0^{\mathrm{S}} = r\sum_{i=1}^{n} \frac{q_i\beta_i}{\mu+\gamma_i+\sigma+f\sigma r\left(\sum_{k=1}^{i-1}\frac{\beta_k}{\gamma_k}+\frac{\beta_i}{\mu+\gamma_i}\right)},$$

where

$$q_i = \prod_{j=1}^{i-1} \frac{\gamma_j}{\mu+\gamma_j+\sigma+f\sigma r\left(\sum_{k=1}^{i-1}\frac{\beta_k}{\gamma_k}+\frac{\beta_i}{\mu+\gamma_i}\right)}.$$

Appendix B. The endemic equilibrium

We now show that when $R_0 > 1$ both the models have a unique non-zero endemic equilibrium and derive a single equation for the equilibrium of each model that can be easily solved numerically.

B.1. The endemic equilibrium for the SCT-DI model

We now show there exists a unique endemic equilibrium when the infection-free equilibrium is unstable ($R_0 > 1$). The endemic equilibrium for (3.1) satisfies the equation:

$$\begin{aligned} p_i\lambda S^* &= (\mu+v_i+\sigma+f\sigma(L_i+M_i))I_i^* = \left(\mu+v_i+\sigma+f\sigma\left(\frac{rT_{\mathrm{M}}I^*}{N^*}+\frac{r\widetilde{T}_i\beta_iS^*}{N^*}\right)\right)I_i^* \\ &= \left(\mu+v_i+\sigma+f\sigma\left(rT_{\mathrm{M}}\left(1-\frac{S^*}{N^*}\right)+r\widetilde{T}_i\beta_i\frac{S^*}{N^*}\right)\right)I_i^* := \left(a_i+b_i\frac{S^*}{N^*}\right)I_i^*, \end{aligned} \tag{B.1}$$

where $a_i = \mu+v_i+\sigma+rf\sigma T_{\mathrm{M}}$, $b_i = f\sigma r(\widetilde{T}_i\beta_i - T_{\mathrm{M}})$. Hence,

$$I_i^* = \frac{p_i\lambda S^*}{a_i+b_i\frac{S^*}{N^*}},$$

which gives

$$\lambda^* = r\sum_{i=1}^{n}\frac{\beta_iI_i^*}{N^*} = r\sum_{i=1}^{n}\frac{\beta_ip_i\lambda^*S^*}{N^*\left(a_i+b_i\frac{S^*}{N^*}\right)}.$$

That is,

$$1 = r\frac{S^*}{N^*}\sum_{i=1}^{n}\frac{\beta_ip_i}{a_i+b_i\frac{S^*}{N^*}}. \tag{B.2}$$

The fraction of the population that is susceptible at the equilibrium as $x := S^*/N^* \in (0,1)$ is used as a variable to define the function

$$H_{\mathrm{D}}(x) := r\sum_{i=1}^{n}\frac{\beta_ip_i}{\frac{a_i}{x}+b_i} - 1, \tag{B.3}$$

where $H_D(x) = 0$ at the equilibrium. Because $H_D(x)$ is an increasing function, $\lim_{x\to 0} H_D(x) = -1$, and

$$\lim_{x\to 1} H_D(x) = r\sum_{i=1}^{n} \frac{\beta_i p_i}{a_i + b_i} = r\sum_{i=1}^{n} \frac{\beta_i p_i}{\mu + \nu_i + \sigma + f\sigma r \widetilde{T}_i \beta_i} = R_0 - 1,$$

there exists a unique solution of $H_D(\hat{x}) = 0$ for $\hat{x} \in (0,1)$, if and only if $R_0 > 1$.

Combining the equilibrium equation for (3.1), $\mu(S^0 - S^*) = \lambda^* S^*$, and (B.1) we have

$$I_i^* = \frac{p_i}{a_i + b_i \hat{x}} \mu(S^0 - S^*). \tag{B.4}$$

Hence,

$$\sum_{i=1}^{n} I_i^* = \mu(S^0 - S^*) \sum_{i=1}^{n} \frac{p_i}{a_i + b_i \hat{x}} = \mu(S^0 - S^*) G(\hat{x}),$$

where $G(\hat{x}) := \sum_{i=1}^{n} \frac{p_i}{a_i + b_i \hat{x}}$.

Define the function $F(\hat{x}) := r\sum_{i=1}^{n} \beta_i p_i/(a_i + b_i\hat{x})$ of the equilibrium solution $\hat{x}$ and note that $N^* = S^* F(\hat{x})$. Therefore,

$$S^* + \mu(S^0 - S^*) G(\hat{x}) = S^* F(\hat{x}), \tag{B.5}$$

or

$$S^* = \frac{\mu G(\hat{x})}{\mu G(\hat{x}) + F(\hat{x}) - 1} S^0. \tag{B.6}$$

From (B.5) it also follows that

$$\mu(S^0 - S^*) = \frac{F(\hat{x}) - 1}{G(\hat{x})} S^*. \tag{B.7}$$

Substituting (B.7) into (B.4) gives

$$I_i^* = \frac{p_i(F(\hat{x}) - 1)}{(a_i + b_i\hat{x}) G(\hat{x})} S^* = \frac{\mu p_i (F(\hat{x}) - 1)}{(a_i + b_i \hat{x})(\mu G(\hat{x}) + F(\hat{x}) - 1)} S^0. \tag{B.8}$$

Because $F(\hat{x}) = 1/\hat{x} > 1$ and (B.5) we can conclude that $S^* > 0$ and $I_i^* > 0$.

B.2. The endemic equilibrium for the SCT-SP model

The equilibrium equations for the SP model (3.2),

$$\lambda S^* = (\gamma_1 + \mu + \sigma + f\sigma(L_1 + M_1)) I_1^*$$
$$\gamma_{i-1} I_{i-1}^* = (\gamma_i + \mu + \sigma + f\sigma(L_i + M_i)) I_i^*, \quad 2 \leqslant i \leqslant n,$$

can be combined to give the conditions

$$\lambda S^* = \left(A_1 + B_1 \frac{S^*}{N^*} \right) I_1^*, \tag{B.9}$$

and

$$I^*_{i-1} = \left(A_i + B_i \frac{S^*}{N^*}\right) I^*_i, \quad i = 1, \ldots, n-1. \tag{B.10}$$

That is,

$$I^*_i = \prod_{j=i+1}^{n} \left(A_j + B_j \frac{S^*}{N^*}\right) I^*_n, \quad i = 1, \cdots, n-1, \tag{B.11}$$

where $A_i = (\gamma_i + \mu + \sigma + f\sigma r T_{\mathrm{M}})/\gamma_{i-1}$, $B_i = f\sigma r(J_{M_i} - T_{\mathrm{M}})/\gamma_{i-1}$, with $\gamma_0 = 1$ and J_{M_i} are given by (4.13).

Substituting λ and (B.11) into (B.9) then leads to

$$r\frac{S^*}{N^*}\left(\sum_{i=1}^{n} \beta_i \prod_{j=i+1}^{n} \left(A_j + B_j \frac{S^*}{N^*}\right)\right) = \prod_{j=1}^{n} \left(A_j + B_j \frac{S^*}{N^*}\right). \tag{B.12}$$

Defining $x := S^*/N^*$ and dividing (B.12) by the right-hand side, we obtain

$$H_{\mathrm{S}}(x) := rx \sum_{i=1}^{n} \frac{\beta_i}{\prod_{j=1}^{i}(A_j + B_j x)} - 1 = 0. \tag{B.13}$$

The two end limits are $\lim_{x\to 0} H_{\mathrm{S}}(x) = -1$ and

$$\lim_{x\to 1} H_{\mathrm{S}}(x) = r \sum_{i=1}^{n} \frac{\beta_i}{\prod_{j=1}^{i}(A_j + B_j)} - 1 = R_0 - 1.$$

Therefore, if $R_0 > 1$, there exists a solution $\tilde{x} \in (0,1)$ of (B.13) and the solution is unique if the derivative of $H_{\mathrm{S}}(x)$ is positive.

The derivative of $H_{\mathrm{S}}(x)$ is given by

$$\begin{aligned} H'_{\mathrm{S}}(x) &= r\left(\sum_{i=1}^{n} \frac{\beta_i}{\prod_{j=1}^{i}(A_j + B_j x)} - x \sum_{i=1}^{n} \frac{\beta_i}{\prod_{j=1}^{i}(A_j + B_j x)} \sum_{j=1}^{i} \frac{B_j}{A_j + B_j x}\right) \\ &= r \sum_{i=1}^{n} \frac{\beta_i}{\prod_{j=1}^{i}(A_j + B_j x)} \left(1 - \sum_{j=1}^{i} \frac{B_j x}{A_j + B_j x}\right). \end{aligned} \tag{B.14}$$

It follows from (3.9)–(3.12) that $J_{M_i} \leqslant T_{\mathrm{M}}$ for all three cases of $J(i)$. Then, $B_j \leqslant 0$ for all j, and $A_j + B_j x > 0$ for $x \in (0,1)$. Hence $H'_S(x) > 0$, which ensures the uniqueness of the endemic equilibrium.

To solve for I^*, note that substituting $\tilde{x}$ into (B.11) yields

$$I^*_i = \prod_{j=i+1}^{n} (A_j + B_j x) I^*_n, \quad i = 1, \ldots, n-1. \tag{B.15}$$

It follows from $S^* = xN^* = x(S^* + \sum_{i=1}^{n} I^*_i)$ that

$$S^* = \frac{\sum_{i=1}^{n} I^*_i}{1/\tilde{x} - 1} = \frac{\sum_{i=1}^{n} \prod_{j=i+1}^{n} \left(A_j + B_j \tilde{x}\right)}{1/\tilde{x} - 1} I^*_n. \tag{B.16}$$

Combining the equilibrium equation (3.2) for S with (B.9) and (B.15) yields

$$\mu(S^0 - S^*) = \lambda S^* = (A_1 + B_1\tilde{x})I_1^* = \prod_{j=1}^{n}\left(A_j + B_j\tilde{x}\right)I_n^*,$$

which when combined with (B.16) gives

$$\mu S^0 = \left(\mu\frac{\sum_{i=1}^{n}\prod_{j=i+1}^{n}\left(A_j + B_j\tilde{x}\right)}{1/\tilde{x} - 1} + \prod_{j=1}^{n}\left(A_j + B_j\tilde{x}\right)\right)I_n^*.$$

Solving this equation for I_n^* gives

$$I_n^* = \frac{\mu S^0}{\left(\mu\frac{\sum_{i=1}^{n}\prod_{j=i+1}^{n}\left(A_j+B_j\tilde{x}\right)}{1/\tilde{x}-1} + \prod_{j=1}^{n}\left(A_j + B_j\tilde{x}\right)\right)}. \tag{B.17}$$

and substituting this into (B.11) and (B.16) then completely solves for S^* and I_i^*, $i = 1, \ldots, n$.

References

[1] American Civil Liberties Union Web Site on AIDS, Available from <http://www.aclu.org/issues/aids/hmaids.html>.
[2] R.M. Anderson, R.M. May, G.F. Medley, A. Johnson, A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV) the causative agent of AIDS, IMA, J. Math. Med. Biol. 3 (1986) 229.
[3] D. Baltimore, Lessons from people with nonprogressive HIV infection, New England J. Med. 332 (1995) 259.
[4] Y. Cao, L. Qin, L. Zhang, J. Safrit, D.D. Ho, Virologic and immunologic characterization of long-term survivors of HIV type 1 infection, New England J. Med. 332 (1995) 201.
[5] Ferry, Barbara, Teleconference debates partner notification of AIDS, The New Mexican, B-3, October 20, 1999.
[6] K.P. Hadeler, C. Castillo-Chavez, A core group model for disease transmission, Math. Biosci. 128 (1995) 41.
[7] D.R. Henrard, J.F. Phillips, L.R. Muenz, W.A. Blattner, D. Weisner, M.E. Eyster, J.J. Goedert, Natural history of HIV-1 cell-free viremia, JAMA 274 (1995) 554.
[8] J.M. Hyman, J. Li, E.A. Stanley, The differential infectivity and staged progression models for the transmission of HIV, Math. Biosci. 155 (1999) 77.
[9] J.M. Hyman, J. Li, E.A. Stanley, The initialization and sensitivity of multigroup models for the transmission of HIV, J. Theor. Biol. 208 (2001) 227.
[10] J.A. Jacquez, C.P. Simon, J. Koopman, L. Sattenspiel, T. Perry, Modelling and analyzing HIV transmission: the effect of contact patterns, Math. Biosci. 92 (1988) 119.
[11] J.A. Jacquez, J.S. Koopman, C.P. Simon, I.M. Longini, Role of the primary infection in epidemics of HIV infection in gay cohorts, J AIDS 7 (1994) 1169.
[12] J.A. Jacquez, C.P. Simon, J. Koopman, Core groups and the R_0s for subgroups in heterogeneous SIS and SI models, in: Mollison (Ed.), Epidemic Models: Their Structure and Relation to Data, Cambridge University, Cambridge, 1995, p. 279.
[13] M. Kretzschmar, Y.T.H.P. vanDuynhoven, A.J. Severijnen, Modeling prevention strategies for gonorrhea and chlamydia using stochastic network simulations, Am. J. Epidem. 144 (1996) 306.
[14] S.E. Landis, V.J. Schoenbach, D.J. Weber, M. Mittal, B. Krishan, K. Lewis, G.G. Koch, Results of a randomized trial of partner notification in cases HIV infection in North Carolina, New England J. Med. 326 (1992) 101.
[15] X. Lin, Qualitative analysis of an HIV transmission model, Math. Biosci. 104 (1991) 111.
[16] X. Lin, H.W. Hethcote, P. Van den Driessche, An epidemiological model for HIV/AIDS with proportional recruitment, Math. Biosci. 118 (1993) 181.

[17] I.M. Longini, W.S. Clark, M. Haber, R. Horsburgh, The stages of HIV infection: waiting times and infection transmission probabilities, in: Castillo-Chavez, Levin, Shoemaker (Eds.), Mathematical Approaches to AIDS Epidemiology, Lecture Notes in Biomathematics, vol. 83, Springer, New York, 1989, p. 111.
[18] B.A. Macke, J.E. Maher, Partner notification in the United States: an evidence based review, Am. J. Prev. Med. 17 (1999) 230.
[19] S. Higginbotham, R. Holmes, H. Stone, J. Beil, S. Costa, S. Paul, Adoption of protective behaviors among persons with recent HIV infection and diagnosis—Alabama, New Jersey, and Tennessee, 1997–1998, MMWR 49 (2000) 512.
[20] J. Muller, M. Kretzschmar, K. Dietz, Contact tracing in stochastic and deterministic epidemic models, Math. Biosci. 164 (2000) 39.
[21] C. Norwood, Mandated life versus mandatory death: New York's disgraceful partner notification record, J. Commun. Health 20 (1995) 161.
[22] T.R. O'Brien, W.A. Blattner, D. Waters, M.E. Eyster, M.W. Hilgartner, A.R. Coher, N. Luban, A. Hatzakis, L.M. Aledort, P.S. Rosenberg, W.J. Milet, B.L. Kroner, J.J. Goedert, Serum HIV-1 RNA levels and time to development of AIDS in the multicenter haemophilia cohort study, JAMA 276 (1996) 105.
[23] Potterat, J. John, Contact tracing's price not its value, Sex. Transm. Dis. 24 (1997) 519.
[24] R.D. Pratt, J.F. Shapiro, N. McKinney, S. Kwok, S.A. Spector, Virologic characterization of primary HIV-1 infection in a health care worker following needle sick injury, J. Infect. Dis. 172 (1995) 851.
[25] T.C. Quinn, Acute primary HIV infection, JAMA 278 (1997) 58.
[26] K.H. Rotherberg, S.J. Paskey, The risk of domestic violence and women with HIV infection: implications for partner notification, public policy, and the law, Am. J. Public Health 85 (1995) 1569.
[27] R.A. Royce, A. Sena, W. Cates, M.S. Cohen, Sexual transmission of HIV, New England J. Med. 336 (1997) 1072.
[28] C.B. Savitch, Why too many are dying of AIDS at the altar of privacy, ACP-ASIM Observer, March 1999.
[29] Vernon, T.M.R.E. Hoffman, R.F. Wykoff, C.W. Heath, G.W. Rutherford, J.M. Woo, Contact tracing to control the spread of HIV, Lett. JAMA 260 (1998) 3274.
[30] M.T. Wong, M.J. Dolan, E. Kozlow, R. Doe, G.P. Melcher, D.S. Burke, R.N. Boswell, M. Vahey, Patterns of virus burden and T cell phenotype are established early and are correlated with the rate of disease progression in hiv type 1 infected persons, J. Infect. Dis. 173 (1996) 877.
[31] R.F. Wykoff, C.W. Heath, S.L. Hollis, S.T. Leonard, C.B. Quiller, J.L. Jones, M. Artzrouni, R.L. Parker, Contact tracing to identify human immunodeficiency virus infection in a rural community, JAMA 259 (1998) 3563.